2020

Determinants of the DNA binding and gene regulatory specificity for type II nuclear receptor signaling

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

DETERMINANTS OF THE DNA BINDING AND GENE REGULATORY SPECIFICITY FOR TYPE II NUCLEAR RECEPTORS

by

ASHLEY PENVOSE

B.S., Northeastern University, 2006
M.A., Boston University, 2016

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2020

Approved by

First Reader _____

Trevor Siggers, Ph.D.
Associate Professor of Biology

Second Reader _____

Ulla Hansen, Ph.D.
Professor of Biology

DEDICATION

For Ryan

ACKNOWLEDGMENTS

First, thank you to my advisor, Trevor Siggers, for allowing me to pursue my interest in the nuclear receptors and for his mentorship throughout the years. Thank you to my committee members, both past and present (Drs. Thomas Gilmore, Ulla Hansen, Chip Celenza, David J. Waxman, and Valentina Perissi), all of whom have guided me in many different ways throughout the years.

To all of my labmates, you helped to create a dynamic and fun environment in which to conduct my research and I am very grateful to have worked with you all. I would particularly like to thank Jessica Keenan, my partner in science, with whom I collaborated closely on this work.

Thank you to the Biology department at Boston University, which is full of wonderful and supportive people serving in so many different capacities; and particularly to Christina Honeycutt for guiding me through the thesis process. A big thank you to GWISE at Boston University, a group of amazing and talented women too numerous to name individually, for the endless inspiration and for shaping me as a leader.

I would also like to thank my many friends who have helped me in so many ways. First, thank you to Bob Patterson for encouraging me to pursue my goals and return to graduate school. I particularly want to thank Kellen Andrilenas and Ethan Apter, both of whom went above and beyond in helping me through a very difficult transition during my graduate career. I am very grateful for the companionship provided by my wonderful graduate cohort, particularly Sandy Serizier, Mark Deehan, Dan Zuch, Alex Bloom, and

Albert Mondragon, who all provided fun and support throughout this process. Thank you to many members of the Biological Design Center, in particular Ali Lashkaripour, Sam Matos, Rita Chen, Krishna Simhadri, Sam Gan, Meghan Bragdon, and Zack Heins, for the support and feedback during the writing process. Also, thank you to my many other friends who have supported me throughout the years including James Ortega, Jose Medrano, Christin Clohosey, Tiffany Dill, Navid Dianati, and Julie Konig. All of you have been amazing friends and I am so glad to have you in my life.

Thank you to my mom and dad for always supporting my educational pursuits and helping me to achieve my dreams

Lastly, a huge thank you to my partner, Luis Ortiz, who enriches my life in countless ways every single day. He has served as a steady source of support, humor, and comfort throughout my graduate career, and particularly during the thesis-writing process. I can’t imagine having made it here without him.

DETERMINANTS OF THE DNA BINDING AND GENE REGULATORY SPECIFICITY FOR TYPE II NUCLEAR RECEPTOR SIGNALING

ASHLEY PENVOSE

Boston University Graduate School of Arts and Science, 2020

Major Professor: Trevor Siggers, Associate Professor of Biology

ABSTRACT

The type II nuclear receptors (NRs) function as heterodimeric transcription factors with the retinoid X receptor (RXR) to regulate diverse biological processes in response to endogenous ligands and therapeutic drugs. Due to their importance as therapeutic targets, the NRs have been extensively studied; however, the rules dictating NR transcriptional specificity remain unclear.

Type II NRs regulate both distinct and overlapping gene programs. DNA-binding specificity has been proposed as a primary mechanism dictating how individual NRs distinguish their genomic targets. However, many in vivo targets of NRs identified in ChIP-seq data lack an identifiable binding site, suggesting that current models of DNA binding specificity of the type II NRs are incomplete. A more thorough characterization of the DNA binding and gene regulatory specificity of the type II NRs will be informative for understanding the role of DNA binding in achieving NR transcriptional specificity. NRs function can be altered by ligand binding, post-

translational modifications (PTMs), and interactions with other proteins. Therefore, to understand how cellular context may alter NR regulatory specificity, analysis of NR DNA binding in the presence of different ligands or cell-specific coregulator proteins will be informative.

I used protein binding microarrays (PBMs) to characterize the DNA binding preferences of twelve NR:RXRα heterodimers. I find more promiscuous NR DNA binding than has been reported, challenging the view that NR binding is defined by half-site spacing. I show that NRs bind DNA using two distinct modes, explaining widespread half-site binding in vivo. Finally, we show that current models of NR DNA binding preferences better reflect binding-site activity rather than binding-site affinity. Examining how DNA binding is altered in a cellular context, I find that NR-DNA binding is significantly altered in the presence of soluble nuclear components, such as other transcription factors and transcriptional coregulators. In the presence of other nuclear proteins NR DNA binding is less promiscuous, suggesting interactions with nuclear proteins can modulate NR DNA binding specificity. Our rich PBM dataset and revised NR binding models provide a framework for understanding NR specificity and will facilitate more accurate analyses of genomic datasets.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

α	alpha
A	adenine
AF-1	activation function 1
AF-2	activation function 2
β	beta
bp	base pairs
C	cytosine
CAR	constitutive androstane receptor
ChIP-seq	chromatin immunoprecipitation followed by sequencing
CoRs	coregulatory proteins
DBD	DNA binding domain
DBDmut	DNA-binding domain mutants
°C	degrees Celsius
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DRs	direct repeats
DTT	dithiothreitol
EMSAs	electrophoretic mobility shift assays
ER	estrogen receptor
ERE	everted repeat element
ERRs	estrogen related receptors
FBS	fetal bovine serum
FXR	farnesoid X receptor
G	guanine
g	gravity
γ	gamma
GR	glucocorticoid receptor
h	hour(s)
HAT	histone acetyltransferase
HDAC	histone deacetylase activity
HelT	helix twist
HMT	histone methyltransferase
HNF4	hepatocyte nuclear factor 4
IRE	inverted repeat
K	either guanine or thymine
K_d	equilibrium dissociation constant

k_{off}	dissociation rate constant
k_{on}	association rate constant
kDa	kilodalton
LB	Luria-Bertani broth
LBD	ligand binding domain
Luc2	luciferase
LXRs	liver x receptors
µl	microliter
µM	micromolar
µm	micron
min	minute(s)
mg	milligram
ml	millilter
MWG	major groove width
nextPBM	nuclear extract protein binding microarray
Nluc	Nanoluc
nM	nanomolar
NR:RXRα	NR heterodimers
NRs	nuclear receptors
nt	nucleotide
PBM	protein binding microarray
PBS	phosphate buffered saline
PFMs	position frequency matrices
pmol	picomolar
PMSF	phenylmethylsulfonyl fluoride
PPAR	peroxisome-proliferator activated receptor
PPARγ-AL	PPARγ from 3T3-L1 adipocyte lysates
ProT	propeller twist
PTMs	post-translational modifications
PWMs	position-weight matrices
PXR	pregnane X receptor
R	either adenine or guanine
RARs	retinoic acid receptors
rpm	rotations per minute
RXR	retinoid X receptor
sec	second(s)
SNV	single-nucleotide variant
SRC1	steroid receptor coactivator 1

T	thymine
TFs	transcription factors
THRs	thyroid hormone receptors
TPhP	triphenyl phosphate
VDR	vitamin D receptor

CHAPTER ONE: INTRODUCTION

1 General overview

Nuclear receptors (NRs) are a family of transcription factors (TFs) that are able to directly sense environmental changes through ligand binding to regulate gene transcription (Everett and Lazar, 2013). NRs are involved in the pathogenesis of a variety of human diseases including type II diabetes, atherosclerosis, and cancer (Tenbaum and Baniahmad, 1997; Castrillo and Tontonoz, 2004). The ability of NRs to directly bind ligand and regulate gene transcription has made them attractive therapeutic targets, and many drugs directly target the NRs (Barish and Evans, 2004). Due to the central role NRs play in human health, it is important that we work toward a deeper understanding of how NRs find their DNA targets and exert their gene regulatory effects.

1.1 Discovery of the type II NRs

The discovery of the NRs is rooted in the field of endocrinology, as experiments seeking to elucidate the mechanisms of hormone and steroid action led to the identification of genes for the estrogen receptor and glucocorticoid receptor (Evans and Mangelsdorf, 2014). Isolation of cDNAs for the estrogen, glucocorticoid, and thyroid hormone receptors revealed evolutionarily conserved sequences, leading researchers to search for other genes following this conserved template (Green et al., 1986; Hollenberg et al., 1985; Sap et al., 1986; Weinberger et al., 1985). NRs were shown to be

conserved throughout metazoan evolution (Markov et al., 2017), with 48 NRs encoded in humans (Evans and Mangelsdorf, 2014).

NRs discovered based on sequence conservation were initially referred to as "orphan receptors", as their ligands were unknown (Willson and Moore, 2002). The ligand-binding domain (LBD) and DNA-binding domain (DBD) of NRs function largely independently of one another, which allowed researchers to swap domains of different NRs in the search for ligands of these orphan NRs (Giguère et al., 1986; Green and Chambon, 1987). Thus, the DNA-binding domain of an NR for which the DNA response element was already known could be coupled to the ligand binding domain of an NR with an unknown ligand (Green and Chambon, 1987). Cells could then be co-transfected with both the cDNA encoding the chimeric NR and a reporter plasmid containing the known DNA response element coupled to a reporter gene. Expression of this reporter gene could then be used to quantify transcription in response to different ligands, allowing for a rapid way to screen a newly discovered NR against a panel of candidate ligands (Giguère et al., 1986; Levin et al., 1992). This reverse endocrinology approach identified 9-*cis* retinoic as the ligand for the retinoid X receptor (RXR) (Heyman et al., 1992; Levin et al., 1992; Mangelsdorf et al., 1990). These screens also revealed that RXR was capable of functioning as a heterodimer with a protein that responded to chemicals known to cause proliferation of peroxisomes, leading to the subsequent identification of another NR, the peroxisome-proliferator activated receptor (PPAR) (Issemann and Green, 1990; Kliewer et al., 1992).

The ability of RXR to function as a heterodimer led to the identification of many other NR partners of RXR, including the liver x receptors (LXRs), retinoic acid receptors (RARs), farnesoid X receptor (FXR), vitamin D receptor (VDR), pregnane X receptor (PXR), constitutive androstane receptor (CAR), and thyroid hormone receptors (THRs) (Fig. 1.1) (Evans and Mangelsdorf, 2014; Kliewer et al., 1992; Zhang et al., 1992). NRs that function as heterodimers with RXR became known as the type II NRs. Many type II NRs are key regulators of metabolic homeostasis and their endogenous ligands are metabolic derivatives, thus directly coupling changes in cellular nutrient availability to regulation of metabolic pathways (Evans and Mangelsdorf, 2014).

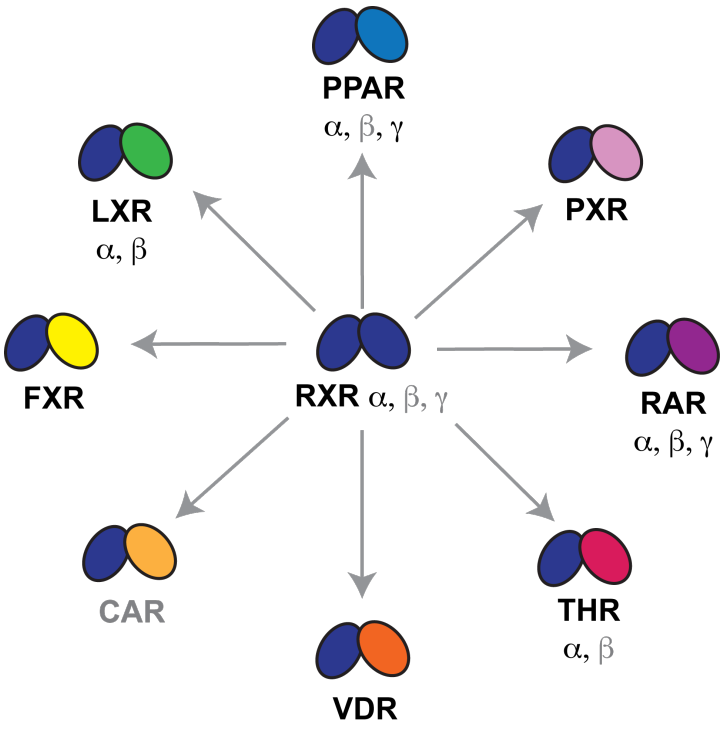


Figure 1.1 Type II nuclear receptors

The type II nuclear receptors that can dimerize with RXRα are shown. All the NRs examined within this thesis are shown in black.

Other type II NRs regulate development (RARs), general homeostasis (VDR), basal metabolic rate (THR), and protection from xenobiotics (CAR and PXR) (Evans and Mangelsdorf, 2014).

1.2 Conserved Structure of Nuclear Receptors

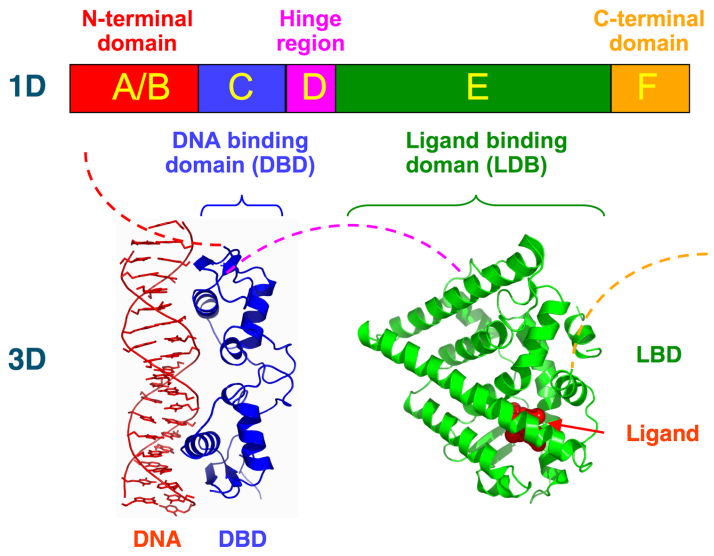


Figure 1.2: Structural organization of the nuclear receptors.

Structural organization of the NRs is shown in a 1D schematic alongside the corresponding 3D structure. Shown are the N-terminal domain containing the activating function-1 region (A/B region), the DNA binding domain (C region), the hinge (D region), ligand binding domain (E region), and C-terminal domain containing the activating function-2 region (F region). Regions shown as dotted lines within the crystal structure are flexible and could not be structurally resolved. Figure obtained from Wikipedia and is in the public domain.

NRs share a modular, conserved structure (Fig. 1.2) (Glass and Ogawa, 2005; Green and Chambon, 1987). NR domains include the N-terminal activation function 1 region (AF-1), which regulates transcription in the absence of ligand; the DBD composed of two C4 zinc fingers that result in sequence-specific binding to DNA; the hinge region,

which serves as a flexible linker connecting the DBD to the LBD; and the C-terminal LBD, which contains the ligand-dependent activation function 2 region (AF-2) (Jin and Li, 2010).

1.3 DNA binding preferences of the type II NRs

NRs rely on sequence-specific DNA binding to regulate target genes. Early experiments seeking to establish the DNA sequences recognized by type II NRs utilized electrophoretic mobility shift assays (EMSAs) to test sequences from promoters of ligand-responsive genes for each NR (IJpenberg et al., 1997; Kurokawa et al., 1993; Meier et al., 1993; Perlmann et al., 1993; Xie et al., 2000). The majority of response elements identified were variations of two direct repeats (DRs) of the sequence 5'-RGKTCA-3' (T, G, C, and A represent the standard DNA bases, R can be either A or G, and K can be either G or T) with a variable length between the two half-sites, often referred to as DR0-5 (Evans and Mangelsdorf, 2014; IJpenberg et al., 1997). The 3-4-5 rule, which stated that DR spacer length is a primary determinant of NR specificity, became the prevailing paradigm for NR target specificity (Evans and Mangelsdorf, 2014; Perlmann et al., 1993; Umesono et al., 1991). However, subsequent experiments revealed that there were exceptions to this simple paradigm, as many NRs share the same spacer preferences, and a single NR can bind to DRs with different spacer lengths (Evans and Mangelsdorf, 2014; Perlmann et al., 1993). Additionally, the DNA sequences directly flanking the DRs have also been demonstrated to impart binding specificity for PPARγ, THR, and RAR (Juge-Aubry et al., 1997; Phan et al., 2010).

The DNA binding preferences of the NRs were turned into position-weight matrices (PWMs), a commonly used mathematical model that describes the log-likelihood of a TF binding to any DNA sequence, which were used to predict NR binding in the genome (Stormo and Zhao, 2010). With the onset of genome-wide occupancy studies, it has become clear that a large portion of NR binding cannot be explained by existing PWMs (Boergesen et al., 2012; Everett and Lazar, 2013). For instance, binding sites as identified by these PWMs are found in only 45-70% of the genomic regions where PPARγ binding is detected in ChIP-seq, suggesting that divergent sequences are capable of mediating binding or that PPARγ is being indirectly recruited to DNA by other factors (Boergesen et al., 2012; Everett and Lazar, 2013; Lefterova et al., 2008; 2010; Nielsen et al., 2008). Attempts to computationally identify divergent binding sites within these existing PPARγ ChIP-seq datasets have yielded a few new candidate sequences, but no validation has been performed for these sites (Lemay and Hwang, 2006). Additionally, even when these new models are included, a large portion of PPARγ binding events remain unexplained (Lemay and Hwang, 2006). These analyses support the idea that divergent binding sites for NRs exist and that utilizing a high-throughput approach to characterize binding, such as the PBM approach utilized in Chapter 2, are necessary for future genomic analyses of the NRs.

1.4 Nuclear receptor function

As the innate ability of NRs to bind ligand has made them attractive therapeutic targets, the ligands and LBDs of NRs have been extensively studied (Choudhary and Malek, 2016; Heemers and Tindall, 2010; Nomiyama and Bruemmer, 2008). Ligand-

binding induces a conformational change in the LBD of NRs that alters the structure of these proteins, allowing for recruitment of coactivators through a conserved LxxLL motif found in a variety of coregulators (McInerney et al., 1998). Corepressors are recruited through a similar LxxH/IIxxxI/L motif that interacts with NRs in a region overlapping the region used to interact with coactivators, thus NR recruitment of coactivators or corepressors is largely mutually exclusive (Hu and Lazar, 1999; Perissi et al., 1999; Perissi and Rosenfeld, 2005).

Initial characterization of NR signaling suggested that in the absence of ligand the type II NRs are pre-bound to DNA and function as repressive TFs, interacting with ATP-dependent chromatin remodelers, histone deacetylases, and histone demethylases to create a repressive chromatin environment (Everett and Lazar, 2013). In this model, ligand binding leads to the exchange of corepressors for coactivators, that in turn promote a permissive chromatin environment resulting in upregulation of transcription (Perissi and Rosenfeld, 2005). While these initial insights into NR function were attractive for their simplicity, recent research has demonstrated that NR function is more nuanced than initially thought, with DNA response element sequence, ligands, and post-translational modifications capable of altering NR function (discussed in detail in sections 1.6 and 1.8).

1.5 Structural perspectives on NR function

Crystal structures of isolated NR DBDs and LBDs gave initial insights into the structural elements governing NR function (Gampe et al., 2000; Schwabe et al., 1993; Uppenberg et al., 1998; Wurtz et al., 1996). PPARγ:RXRα was the first full-length NR crystal structure to be obtained and revealed extensive cross-domain interactions that impact NR function, which could not be appreciated in earlier structures of isolated NR domains (Chandra et al., 2008). This full-length PPARγ:RXRα crystal structure, obtained in the presence of both DNA and coactivator peptides, gave insight into the different domains of the heterodimer that are involved in DNA recognition (Chandra et al., 2008) It demonstrated that the first zinc finger of each receptor's DBD interacts extensively with the major groove of the DNA response element. Interestingly, it revealed that domains outside of the DBDs also interact directly with DNA (Fig. 1.3). The 5' flanking region of the first half-site of the DR is contacted extensively by the hinge region of PPARγ, which may explain the polarity observed in PPARγ:RXRα binding. In all instances where it has been explored, PPARγ occupies the 5' half-site, and the sequence of the 5' flanking region has been shown to influence

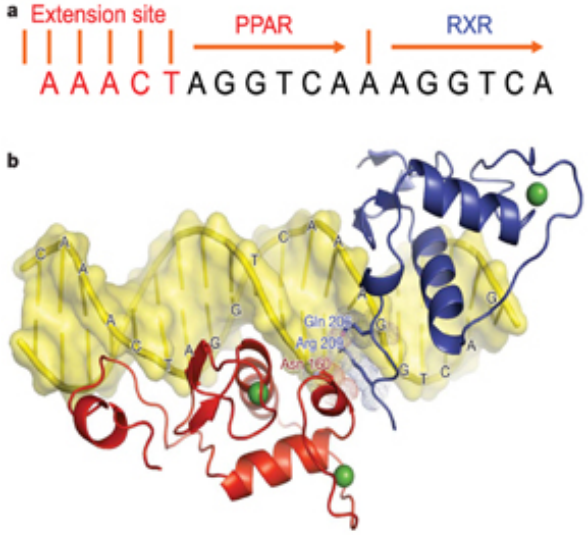


Figure 1.3 Crystal structure of DBDs of PPARγ:RXRα bound to DNA.

a DNA sequence of DR is shown, including the 5' extension site. **b** Crystal structure of PPARγ:RXRα DBDs bound to DNA probe from **a** is shown. Figure adapted from (Chandra et al., 2008).

binding affinity (IJpenberg et al., 1997). This observed polarity is in contrast to other RXR heterodimers, wherein RXR occupies the 5' half-site (Zechel et al., 1994). Additionally, the crystal structure revealed that the DBD-DBD interaction gained in the presence of DNA are facilitated by contact with the spacing nucleotide that shield the hydrophobic amino acids of the contact from interaction with solvent (Chandra et al., 2008).

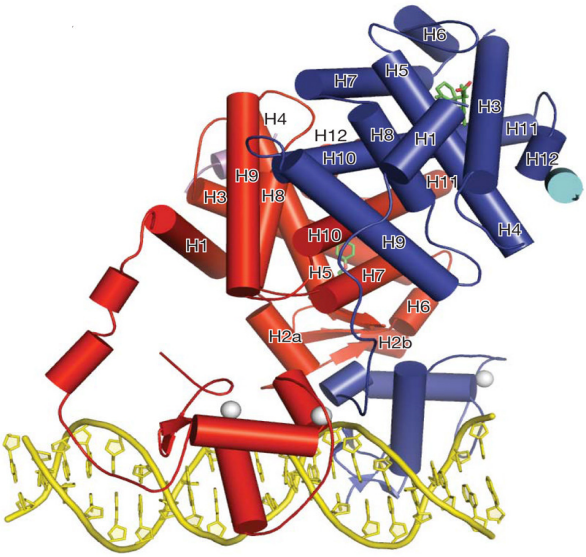


Figure 1.4 Crystal structure of full-length PPARγ:RXRα bound to DNA.

Figure adapted from (Chandra et al., 2008).

The PPARγ:RXRα crystal structure was also elucidating because it revealed how structural changes in one domain of the heterodimer can induce structural changes that propagate to other domains (Fig. 1.4) (Chandra et al., 2008). In the absence of both ligand and DNA, heterodimerization of PPARγ:RXRα occurs only through the LBD. Upon DNA binding, two new cross-receptor interactions are made including a DBD-DBD contact and a contact between the LBD of PPARγ and the C-terminal extension of the DBD of RXRα. This DNA-induced alteration in NR heterodimer structure was further supported by studies of PPAR using hydrogen/deuterium exchange coupled to mass spectrometry and small-angle x-ray scattering (Bernardes et al., 2012), as well as by subsequent full-length crystal structures for VDR:RXR and HNF4α (Lou et

al., 2014). These extensive cross-domain interactions taken together with the conserved structure of the nuclear receptor family may help explain experimental observations that the DNA binding site sequence can allosterically regulate NR function, by altering the structure of the NR and impacting coregulator assembly (discussed in detail in sections 1.6 and 1.8) (Hall et al., 2002; Meijsing et al., 2009).

1.6 DNA sequence as an allosteric regulator of NR function

In the classical model of NR function, unliganded NRs act as transcriptional repressors, and ligand binding acts as a molecular switch allowing NRs to become transcriptional activators (Everett and Lazar, 2013). However, recent high-throughput techniques such as RNA-seq and massively parallel reporter assays have shown that while this model holds true at many loci, liganded NRs also mediate other types of transcriptional regulation, including repression (Everett and Lazar, 2013; Schöne et al., 2016; 2018).

Part of the locus-specific function of NRs may be attributed to the significant structural changes that occur upon binding to different DNA sequences, which alter the exposed surfaces available for interaction with coregulators (Bernardes et al., 2012; Chandra et al., 2008; Lou et al., 2014). Studies of the estrogen receptor (ER), glucocorticoid receptor (GR), and VDR have demonstrated that response element sequence allosterically regulates NR function by influencing the composition of complexes recruited to different DNA sequences (Koszewski et al., 2000; Meijsing et al., 2009; Schöne et al., 2016). Negative response elements that lead to ligand-induced

transcriptional repression have been identified for GR and VDR (Koszewski et al., 2000; Meijsing et al., 2009; Surjit et al., 2011), and both receptors make different DNA contacts on these negative response elements (Koszewski et al., 2000; Meijsing et al., 2009).

While the ability of DNA sequence to allosterically regulate NR function has been studied at small-scale for some NRs, most NRs have not yet been examined and high-throughput approaches are underutilized. The finding that many NRs are allosterically regulated by DNA, coupled with the conserved nature of NR structure, necessitates the study of allosteric regulation of NRs that is described in Chapter 3.

1.7 Nuclear receptors coregulators

Coregulators are key mediators of transcriptional regulation for nearly all TFs, including the NRs (Lonard and O'Malley, 2007). Over 400 coregulators of NRs have been identified (Dasgupta et al., 2014; Lonard and O'Malley, 2007). Coregulators possess a variety of functions with unique regulatory consequences, including histone deacetylase activity (HDAC), histone acetyltransferase (HAT) activity, histone methyltransferase (HMT) activity, methyltransferase activity, demethylase activity, and ATP-dependent chromatin remodeling activity, all of which impact chromatin state and gene activity (Dasgupta et al., 2014; Lonard and O'Malley, 2007; Millard et al., 2013). Coregulators, including the steroid receptor coactivators (SRCs), also serve as scaffold proteins to allow for recruitment and assembly of larger transcriptional complexes, including recruitment of the mediator complex and other core transcriptional machinery to impact many stages of transcription (Viswakarma et al., 2010).

Coregulators are classified as either coactivators or corepressors based on their ability to activate or repress transcription, respectively. Modulation of coregulator activity is partially dictated by post-translational modifications (PTMs), including acetylation, phosphorylation, sumoylation, and ubiquitination (Millard et al., 2013). Additionally, coregulator activity is complex and a single coregulator can possess both coactivator and corepressor functions, with the switch between these functions often being regulated by PTMs (Poulard et al., 2017; Purcell et al., 2011). These PTMs serve as a regulatory interface allowing for integration of upstream signaling cascades, which alter the function of coregulators ultimately leading to changes in transcriptional regulation. Additionally, differential expression of coregulators influences cell-type specific NR function (Dasgupta et al., 2014).

Considering their diverse capabilities, assembly of different coactivator complexes results in differential gene regulation at distinct NR target genes, necessitating an understanding of the molecular determinants of site-specific recruitment of coactivators; this will be carried out in Chapter 3.

1.8 PPARγ and Adipocytes:

PPARγ is highly expressed in adipocytes and is central to adipocyte function, as it is necessary for maintenance of the adipocyte phenotype and is sufficient to convert fibroblast-like pre-adipocytes into mature adipocytes (Barak et al., 1999; Rosen et al., 1999). Mice deficient in PPARγ lack adipocytes (Barak et al., 1999; Rosen et al., 1999). There are three main categories of adipocytes: white adipocytes that store fat; brown

adipocytes that are thermogenic; and brite/beige adipocytes that have an intermediate phenotype between white and brown adipocytes (Giralt and Villarroya, 2013). Differences in function between white, brown, and brite/beige adipocytes are partially due to differential PPARγ-mediated transcriptional programs (Villanueva et al., 2013).

Ligands are key modulators of PPARγ function in adipocytes, impacting both protein conformation and transcriptional regulation (Chrisman et al., 2018). Changes in PPARγ conformation alter the surfaces of the protein that are available for PTMs and coregulator assembly (Chrisman et al., 2018; Villanueva et al., 2013). In turn, PTMs impact the set of genes that are regulated by PPARγ (Choi et al., 2010; 2011; Farmer, 2008). For instance, Sirt1-mediated acetylation of PPARγ (Qiang et al., 2012) and cdk5-mediated phosphorylation of PPARγ (Choi et al., 2011) lead to regulation of distinct gene programs. The ligand-dependent alteration in PPARγ PTMs and subsequent impact on transcriptional regulation have been clearly demonstrated; however, it remains unclear why transcription is only impacted by at specific loci in the genome.

Exogenous ligands of PPARγ have distinct impacts on metabolic function. Environmental pollutants can also act as ligands for the NRs, including PPARγ (Hurst and Waxman, 2003; Maloney and Waxman, 1999) and are often endocrine disruptors that impair human health (Balaguer et al., 2017; Delfosse et al., 2014). Triphenyl phosphate (TPhP), an organophosphate often used as a flame retardant treatment for furniture, was recently identified as an environmental ligand of PPARγ (Wang et al., 2018a). Unlike rosiglitazone, TPhP has an adverse effect on insulin sensitivity in male mice, leading to impaired glucose homeostasis (Wang et al., 2018a). In both bone marrow stromal cells

and 3T3-L1 pre-adipocytes, exposure to TPhP leads to increased adipogenesis (Pillai et al., 2014; Tung et al., 2017). Ongoing studies are investigating the mechanisms underlying the distinct transcriptional programs elicited by environmental ligands of the NRs. In Chapter 3, we examine the DNA binding of NRs and coactivator recruitment from 3T3-L1 adipocyte nuclear lysate. Given the many vital cellular processes that PPARγ regulates, it is essential that we work toward understanding locus-specific function of PPARγ. Due to the central role of PPARγ in adipocytes, they are an ideal cell type in which to carry out these analyses.

1.9 Thesis Rationale:

Our current understanding of NR-DNA binding specificity is lacking, as current models do not distinguish the unique repertoire of target genes selected by NRs, and binding to many in vivo targets remains poorly explained by current models of DNA binding. Additionally, locus-specific function of NRs remains poorly understood.

In Chapter 2 of this thesis, we describe the use of custom-designed protein binding microarrays (PBMs) to comprehensively characterize the DNA binding preferences of twelve different type II nuclear receptors. Our analyses do not support the established target DNA site spacing preferences for type II NRs. We demonstrate that the canonically reported spacing preferences describe functional NR binding rather than binding affinity of purified NRs. Our modeling of NR binding motifs, based on a single-nucleotide variant (SNV) approach, also reveals that functional NR binding may be partially determined by the binding mode utilized by NRs.

In Chapter 3 of this thesis, we use the nuclear extract protein binding microarray (nextPBM) platform to examine (1) the DNA binding specificity of NRs in a cellular context and (2) co-regulator recruitment from nuclear lysates. We find that the DNA binding preferences of PPARγ are changed in a cellular context. We also demonstrate that steroid receptor coactivator 1 (SRC1), a NR coregulator, shows strong DNA sequence-dependent recruitment.

CHAPTER TWO: DNA BINDING SPECIFICITY OF PURIFIED TYPE II NRs

The work presented in this chapter was performed by me and Jessica Keenan, resulting in a co-first author publication (Penvose et al., 2019). All experimental work (cloning, protein expression and purification, PBMs, and reporter experiments) within this chapter were performed by me. The seed sequences used on the custom PBM used for the experiments in this chapter were chosen by me. Jessica Keenan performed the computational portions to create the fully expanded probe set for this PBM design. Author contributions to data acquisition and analyses performed in this chapter are made in the figure legends.

2.1 Abstract

The type II nuclear receptors (NRs) function as heterodimeric transcription factors with the retinoid X receptor (RXR) to regulate diverse biological processes in response to endogenous ligands and therapeutic drugs. DNA-binding specificity has been proposed as a primary mechanism for NR gene regulatory specificity. In this chapter, I used protein-binding microarrays (PBMs) to comprehensively analyze the DNA binding of 12 NR:RXRα dimers. I found more promiscuous NR-DNA binding than has been reported, challenging the view that NR binding specificity is defined solely by half-site spacing. I show that NRs bind DNA using two distinct modes, explaining widespread NR binding to half-sites in vivo. Finally, using reporter assays that I performed and genomic enrichment analyses performed by Jessica Keenan using the models I generated from the PBM experiments, we demonstrate that the current models of NR specificity better reflect

binding-site activity rather than binding-site affinity. Our rich dataset and revised NR binding models provide a framework for understanding NR regulatory specificity and will facilitate more accurate analyses of genomic datasets.

2.2 Introduction.

The type II nuclear receptors (hereafter simply NRs) are ligand-activated transcription factors (TFs) that control diverse cellular processes including development, metabolism, and inflammation (de Aguiar Vallim et al., 2013; Evans and Mangelsdorf, 2014). NRs include peroxisome-proliferator activated receptor (PPAR), liver x receptor (LXR), retinoic acid receptor (RAR), farnesoid x receptor (FXR), pregnane x receptor (PXR), thyroid hormone receptor (THR), and vitamin D receptor (VDR) (Evans and Mangelsdorf, 2014; Kliewer et al., 1999). NRs function as heterodimers with the common partner, the retinoid x receptor (RXR). Individual NR heterodimers can regulate distinct gene programs (Calkin and Tontonoz, 2012); however, the current models of NR-DNA binding specificity are insufficient to fully explain NR-specific gene regulation.

NRs bind the sequence 5'-RGKTCA-3' organized as direct repeats with a variable length spacer of 0–5 base pairs (bp) (DR0-DR5, Fig. 2.1a) (Claessens and Gewirth, 2004; Cotnoir-White et al., 2011; Weikum et al., 2018). Current models propose that DR spacer length is a key determinant of DNA-binding specificity for NRs (Cotnoir-White et al., 2011; Evans and Mangelsdorf, 2014; Kurokawa et al., 1993; Mader et al., 1993; Perlmann et al., 1993). For example, PPAR:RXR dimers prefer binding to DR1 elements, whereas LXR:RXR dimers prefer DR4 elements (Fig. 2.1b). However, there are more

NRs than available spacer lengths; therefore, either DRs are bound by multiple NRs, which presents a problem for achieving NR-specific gene activation, or there are additional determinants of NR-binding specificity beyond DR spacer length.

Differences in DNA-binding specificity for each NR would provide a mechanism for NRs to regulate distinct target genes in vivo. Genome-wide chromatin immunoprecipitation followed by sequencing (ChIP-seq) studies have confirmed known NR preferences for particular DR spacer lengths, and have reinforced the connection between in vitro and in vivo binding (Boergesen et al., 2012; Chatagnon et al., 2015; Le Zhan et al., 2014; Lefterova et al., 2010; Rastinejad et al., 2013; Savic et al., 2016; Soccio et al., 2015). However, these studies have also revealed limitations to current models of NR-DNA binding. For example, PPARγ and LXRα regulate distinct yet overlapping gene programs but do not share a DR element to explain their many common genomic targets (Boergesen et al., 2012; Savic et al., 2016). Additionally, many genomic regions that are bound in vivo lack an identifiable binding site for the NR being investigated (e.g., 90–96% for PPARγ and LXR) (Boergesen et al., 2012). Together, these observations suggest that current models of NR-DNA-binding specificity are incomplete.

To address the need for revised models of NR binding, we use protein-binding microarrays (PBMs) to compare the binding of 12 NR:RXRα dimers to thousands of DNA sequences. To examine DR spacer preferences, we assay NR binding at all spacer lengths (DR0-DR5). We identify both NR-shared and NR-specific binding features in our dataset, and discuss implications for NR-signaling specificity. By integrating PBM and ChIP-seq datasets, we examine the relationship between in vitro and in vivo binding. We

address the role of activity versus affinity in current models of NR specificity by integrating PBM data with reporter gene experiments. Our results demonstrate the limitations of DR spacer length for defining NR specificity and of DNA binding affinity for predicting functional binding events.

2.3 Results

2.3.1 Characterizing NR heterodimer binding with PBMs

I used PBMs to characterize the DNA binding of 12 distinct RXR heterodimers (hereafter NRs). PBMs are double-stranded DNA microarrays that enable the high-throughput study of protein-DNA binding (Berger et al., 2006a). To characterize both DNA-base and DR-spacing preferences, I measured NR binding to over 1600 unique sequences at each of six DR spacer lengths (DR0-DR5). For each DR spacer length, I measured NR binding to 24 starting sequences, which I refer to as seed sequences (Fig. 2.1c). Seed sequences were generated by combining different half-site sequences exhibiting a range of degeneracy from the consensus 5'-RGKTCA-3' (seed sequence selection is detailed in section 2.5.2). Most seed sequences contain two distinct half-site sequences. To assay NR binding specificity for each seed sequence I also measured binding to all possible single-nucleotide variants (SNVs), with each SNV included as a separate probe on the PBM (Fig. 2.1c). This SNV-based approach allows us to generate a binding logo (i.e., energy matrix or position-weight matrix (PWM)) for each individual seed sequence by measuring the impact on binding caused by perturbation at each base position (Fig. 2.1c, Methods). To capture binding preferences for DR spacer and flank sequences, we included SNVs across the spacer sequence and for the five nucleotides

upstream and downstream of the DR. Using this comprehensive SNV-type PBM design, we characterized the DNA-base and DR-spacing preferences of the NRs.

PBM experiments for NR heterodimers (NR:RXRα) were performed by combining purified RXRα with purified samples of each partner NR. Hereafter, we refer to NR:RXRα heterodimers simply by the NR partner, and RXRα:RXRα homodimers as RXRα, unless otherwise stated. Most NRs do not bind DNA with high affinity as homodimers; therefore, proteins were combined at a 3:1 NR:RXRα ratio to force RXRα heterodimer formation (exceptions indicated in Table 2.1). To ensure heterodimer binding, I required that the binding results agreed when performed using antibodies for both RXRα and the non-RXRα partner. Binding profiles using separate antibodies showed strong correlation, demonstrating that both protein partners were bound to each DNA probe at similar levels (Fig. 2.1d, R^2 of antibody replicates in Table 2.1). Binding of homodimers were not correlated with each other, nor with the heterodimers (Supplementary Fig. 2.1), further demonstrating heterodimer binding. To quantify binding specificity, PBM fluorescence values were converted into z-scores using a set of 500 random genomic background sequences (Fig. 2.1e). Validated PPARγ binding sites score significantly above background, down to a z-score of 1.5 (Fig. 2.1e). We set a more stringent z-score cutoff of 3.0 to define the affinity cutoff for functional binding sites. A DR1 DNA binding logo generated for PPARγ agrees well with known logos from ChIP-seq (Fig. 2.1f), demonstrating the sensitivity of our assay. To validate our PBM results with an orthogonal approach, I performed competition electrophoretic-mobility shift assays (EMSAs) to measure the relative binding affinity of PPARγ:RXRα to DNA

sequences bound over a wide range of PBM z-scores (Fig. 2.1g, Supplementary Fig. 2.2). I found strong agreement between the relative binding affinities derived using both approaches ($R^2 = 0.93$). The protein samples used for these experiments were produced in bacterial or insect cells; however, our ability to capture known NR-binding specificity suggests our data reflect native mammalian dimer-binding specificity. These results demonstrate that our PBMs accurately capture sequence-specific binding of NR heterodimers.

2.3.2 NRs bind promiscuously to most DR spacings

To understand NR-signaling specificity, studies have examined the DNA-binding differences between NRs (summarized in Fig. 2.1b, Table 2.2 and Supplementary Data 3 from (Penvose et al., 2019)) (Cotnoir-White et al., 2011; Evans and Mangelsdorf, 2014; Juge-Aubry et al., 1997; Mader et al., 1993; Näär et al., 1991). A prevailing view is that NRs are distinguished by their preference for DR sites with specific half-site spacing) (Cotnoir-White et al., 2011; Evans and Mangelsdorf, 2014; Juge-Aubry et al., 1997; Mader et al., 1993; Näär et al., 1991; Perlmann et al., 1993; Zechel et al., 1994).; however, individual NRs are functional on DR sites with various spacings (Katz et al., 1995; Kurokawa et al., 1993; Miyamoto et al., 1997)·. Therefore, for each NR we examined which DR spacings were bound with sufficient affinity such that they might be functional in vivo.

To visualize the NR-binding landscape, I generated a DNA-binding logo from high-scoring seeds at each DR spacing (Fig. 2.2). Strikingly, for all NRs I was to able generate DNA-binding logos at nearly every DR spacing, demonstrating broader binding

preferences than previously reported. While previous reports have shown strong preferences for DRs of certain spacer lengths, the ability to generate logos at alternative spacer lengths suggests that binding can occur at other spacer lengths (discussed in more detail in 2.3.5). Comparing our logos with published DR binding preferences (Fig. 2.1b), I found high-affinity binding for many NRs at new DR spacings. The binding logos for all NRs exhibit the canonical 5'-RGKTCA-3' sequence preferences in each half-site and agree with base preferences reported by other methods (Isakova et al., 2017; Kulakovskiy et al., 2013). The logo similarity demonstrates broad conservation in NR-binding specificity; however, NR-specific preferences are also present. For example, PPARγ prefers an AT-rich sequence 5' of the first half-site of a DR (Juge-Aubry et al., 1997) and my PPARγ logo shows this extended footprint (Figs. 2.1f, 2.2). Overall, my data reveals that all NR heterodimers can bind to sites with variable DR spacings and with highly overlapping base specificities.

2.3.3 All type II NRs can bind DNA using a half-site mode

I found that all NRs can bind with high affinity to half-sites (Fig. 2.2, final two columns). For all NRs, I obtain both 5'- and 3'-half-site logos, with the exception of PPARγ for which I only find clear 5'-half-site binding (Fig. 2.2). Half-site logos indicate that NR binding is only perturbed by SNVs introduced into one half-site of a DR, and that the SNVs introduced in the other half-site cause no changes in affinity. To illustrate, I show the impact of SNVs on LXRα heterodimer binding to seed sequences with different binding modes (Fig. 2.3a-c, Supplementary Fig.2.3). Critically, my data agrees for PBMs probed with antibodies against either dimer member; therefore, half-sites are

bound by NR heterodimers and are not a result of monomer binding. The presence of both full-site and half-site logos suggests that NRs can engage with DNA in two binding modes: (1) full-site mode where the NR engages with both half-sites and (2) half-site mode where the NR engages with a single half-site (either 5' or 3') (Fig. 2.3a).

To ensure that the widespread half-site binding was not a result of our methodology, we performed several analyses. First, we tested whether half-site binding was due to the orientation of the NR-binding site within the PBM probe with respect to the microarray slide. We find that regardless of orientation of the probe, binding mode is maintained (Supplementary Fig. 2.4). Second, I performed PBM experiments at successively lower concentrations to test whether half-site binding is affected by protein concentration and find nearly identical DNA binding logos at all concentrations (Supplementary Fig. 2.5). Finally, I used EMSA experiments to test the impact of base mutations on a DNA site bound in half-site mode (Fig. 2.1g, sequences P3, P3 5'-Abl, P3 3'-Abl). Critically, the 5' half-site mode of PPARγ:RXRα determined by PBM is corroborated by EMSA experiments (i.e., 5' half-site ablation greatly reduced binding whereas 3' half-site ablation only modestly affected binding) (Fig. 2.1g, Supplementary Fig. 2.2). These results demonstrate that PBM-derived binding modes accurately represent native NR-binding modes.

NRs are known to bind half-sites (Fig. 2.1b), though half-sites have primarily been identified in ChIP-seq data and not through direct binding assays. My analysis clarifies that NR heterodimers can bind half-sites, and can engage in a half-site mode even on canonical DR sites composed of two good half-sites (i.e., both half-sites score

well using PWMs). For example, logos generated for a near-consensus DR1 seed sequence that scores highly by DR1 PWMs reveal both full- and half-site binding modes (Fig. 2.3d). While all NRs bind this site with high-affinity (z-scores are shown), only PPARγ binds in a full-site mode, while other NRs bind in half-site modes that are nearly identical to one another. This shows that binding mode can vary for different NRs on the same DNA site, and that throughout the genome NR-binding to DR sites may in fact be mediated through a half-site binding mode.

2.3.4 Role of monomers in half-site binding

To examine the contribution of each protein within an NR heterodimer to DNA binding, I created DNA-binding domain mutants (DBDmut) of RXRα and PPARγ. Two residues within zinc finger 1 of RXRα and PPARγ that make base-specific contact with DNA were mutated to alanines (K156A and R161A; and K132A and R137A, respectively, Fig. 2.3e) (Chandra et al., 2008). Binding of PPARγ:RXRα-DBDmut is highly correlated using either anti-RXRα or anti-PPARγ antibodies (R^2 of antibody replicates given in Table 2.1), showing that all DNA sites are bound by the mutant as a heterodimer. For PPARγ-DBDmut:RXRα, PBMs performed using an anti-RXR antibody are dominated by RXR homodimer signal, therefore binding of PPARγ-DBDmut:RXRα was determined using only the anti-PPARγ antibody. RXRα homodimer binding was not observed in wild-type heterodimer experiments (see above). All DBD mutant proteins were produced by IVT and PBM data for IVT-produced wild-type dimers agree with experiments using purified proteins, demonstrating that IVT proteins form heterodimers

and function in DNA-binding assays similarly to purified proteins (model curations can be found in Supplementary Data 4 of (Penvose et al., 2019).

To confirm that these mutations abrogated DNA interactions, I examined the binding of mutant homodimers using PBMs. The mutant RXRα (RXRα-DBDmut) bound no sequences with z-score > 3.0 (as compared to a max z-score of 7.0 for PPARγ:RXRα-DBDmut described below), demonstrating an abrogation of sequence-specific DNA binding. The mutant PPARγ (PPARγ-DBDmut) showed binding with z-score > 3.0 to only five seed sequences. Previous experiments have shown residual DNA-binding activity for PPARγ DBD mutants (Temple et al., 2005); therefore, I chose to disregard these five sequences from further analysis of the PPARγ-DBDmut:RXRα heterodimer experiments.

I first examined mutant heterodimer binding to sequences that PPARγ:RXRα binds in full-site mode. As expected, binding in full-site mode was almost completely abrogated for the PPARγ:RXRα-DBDmut (38/39 full-sites were lost). Of these sites, 40% (15/38) were now bound in the 5' half-site mode (e.g., Fig. 2.3f , curation of modes can also be found in Supplementary Data 5 of (Penvose et al., 2019), demonstrating an altered binding mode for the PPARγ:RXRα-DBDmut heterodimer. The remaining 60% (23/38) of these sites were bound with low affinity by PPARγ:RXRα-DBDmut, and scored below our z-score threshold for modeling interactions. The reciprocal mutant experiment with PPARγ-DBDmut:RXRα showed a complete loss of binding (i.e., z-score < 3.0) to nearly all of the full-sites (35/36, note that we have disregarded three sequences in this category as described above, model curations can also be found in

Supplementary Data 5 of (Penvose et al., 2019). These results demonstrate that DNA must be engaged by both dimer partners in order for PPARγ:RXRα to utilize a full-site binding mode, and shows that half-site binding can occur when only one partner can bind DNA.

Next, I examined which partner of the wild-type PPARγ:RXRα dimer engages with DNA when binding in a half-site mode. Of the 34 sequences that PPARγ:RXRα bound in a half-site mode, 53% (18/34) remained bound in half-site mode by PPARγ:RXRα-DBDmut, demonstrating that for these sequences PPARγ is making base-specific contacts with the DNA and can tolerate loss of base-specific DNA contacts mediated by RXRα (Fig. 2.3f). For the remaining 47% (16/34) of sequences bound by PPARγ:RXRα in a half-site mode, the mutant dimer binding was too low affinity to model (i.e., z-score < 3.0). Interestingly, PPARγ-DBDmut:RXRα showed a loss of binding to 82% (29/32, note two sequences in this category were disregarded due to residual PPARγ-DBDmut binding as detailed above) of the half-site sequences. These results demonstrate that a single partner of an NR heterodimer can mediate half-site binding; however, for other sites, mutation of either NR partner can lead to loss of heterodimer binding. The strong impact of mutations to either member of the heterodimer may be attributable to the ability of either partner to engage with the half-site, or to a contribution in binding energy through non-specific interactions from the non-engaged partner, which were abrogated by the mutations we made.

2.3.5 NR spacer preferences do not define high-affinity binding

Previous studies have examined the impact of DR spacer length on NR binding (Cotnoir-White et al., 2011; Evans and Mangelsdorf, 2014; Kurokawa et al., 1993; Perlmann et al., 1993); however, my results demonstrate that NRs can bind in a half-site mode even on DR sites, which complicates the interpretation of these experiments as much of the binding examined may have been occurring to a single half-site and not to the full DR sequence. SNV binding models are advantageous as they allow examination of NR-binding mode on each sequence, thus facilitating a more rigorous assessment of NR spacer preferences. I analyzed the NR-binding landscape to all 24 seed sequences at each DR spacing and used the resulting binding logos to annotate whether each sequence was bound in a full-site or half-site mode (Fig. 2.4).

In contrast to the prevailing view of NR spacer preferences (Cotnoir-White et al., 2011; Evans and Mangelsdorf, 2014; Näär et al., 1991; Rastinejad et al., 1995), I found that NRs can bind with high affinity to DRs at all spacer lengths (Fig. 2.4). For most NRs, high-affinity binding to many DR spacer lengths is predominantly mediated via a half-site binding mode (Fig. 2.4 gray dots). Despite this promiscuous NR binding, my results recapitulate literature-reported NR spacer preferences, which are demonstrated by an enrichment of full-site binding mode and higher z-scores for specific DR spacer lengths (Fig. 2.4 blue dots). For example, PPARγ engages with DR1 sequences almost entirely via a full-site binding mode. Similar observations corroborate previously described DR-spacing preferences, for example LXRs (DR1 & DR4), THRα (DR4), and VDR (DR3) (see Fig. 2.1b). However, for most NRs the increase in binding affinity to

certain DR spacers is more modest than observed for PPARγ, suggesting that spacer preferences do not define the DNA binding landscape of each NR. In fact, for some NRs the canonical DR-spacing preferences appear primarily as enrichment in full-site binding mode, but not a large increase in binding affinity. For example, PPARα preferentially engages with DR1 sites in a full-site binding mode but only binds with moderately higher z-scores to these sites. My results reveal a complicated NR-DNA binding landscape in which DR spacer preferences contribute to altered NR-binding modes and binding affinity, but spacer preferences do not strongly define the landscape of all possible high-affinity binding.

2.3.6 Diverse mechanisms contribute to NR-DNA binding

Despite broad similarities seen in binding logos (Fig.2.2), our dataset reveals that NR-binding differences result from multiple mechanisms: DR-spacing preferences, DNA-base preferences, and DNA-binding-mode differences. To illustrate the roles of spacing preferences and binding modes, we compared the binding of PPARγ and LXRα to DR1 and DR4 sites and observed both NR-shared and NR-specific binding sites (Fig. 2.5a). The LXRα preference for DR4 sites and PPARγ preference for DR1 sites are demonstrated as biases in the z-score distributions. However, as we see high-affinity binding of PPARγ to DR4 sites and LXRα to DR1 sites, the aforementioned preferences do not explain all high-affinity binding. To explicitly test the impact of DR spacing, we examined binding to pairs of seed sequences that differ only in their spacer length (e.g., Fig. 2.5a, sequences DR1.1 and DR4.1). Critically, we examined the DNA-binding mode for each interaction using the DNA-binding logos generated for each seed sequence (Fig.

2.5b). For PPARγ, DR4 sites are bound with lower affinity than corresponding DR1 variants (; however, DR4.1 is still bound with high affinity via a half-site binding mode (Fig. 2.5a,b). In contrast, when LXRα binds via a full-site mode the DR4 variant is bound with higher affinity (DR1.1 vs DR4.1), but when binding via a half-site mode the DR4 variant is bound with lower affinity (DR1.2 vs DR4.2) (Fig. 2.5a). Therefore, both NRs can bind the same sequence with high affinity, but may utilize distinct binding modes. Taken together, these results demonstrate that both spacer preference and binding mode contribute to binding specificity.

To investigate the plasticity of DR spacer preferences, we compared PXR and VDR, which exhibit broadly similar binding to DR1 and DR4 sites but differ for DR3 binding. PXR and VDR bind with nearly identical specificity to DR1 and DR4 sites (Fig. 2.5c, $R^2 = 0.98$ for both); however, the VDR preference for DR3 sites is seen as an increase in z-score for most DR3 sequences (Fig. 2.5d). This example illustrates that NRs can bind similarly on one DR spacing while having distinct binding preferences for another DR spacing.

Next, we asked whether shared spacer preferences might constrain DNA-base preferences. PXR and LXRα both exhibit preferences for DR1 and DR4 sites (Fig. 2.4); their binding profiles are highly correlated for DR1 sites ($R^2 = 0.95$), but show lower correlation on DR4 sites ($R^2 = 0.83$) (Fig. 2.5e). Analysis of the standard DNA-binding logos did not reveal a strong basis for differential DNA-base preferences. However, by directly examining the impact of SNVs on binding via visualization as an energy matrix (which indicates both favorable and unfavorable interactions), we see strong differences

between PXR and LXRα at positions 10 and 12 (Fig. 2.5f). The majority of the PXR-specific binding sites are explained by the existence of a guanine base at position 10 that is highly disfavored by LXRα (G10 carries a z-score penalty of −3.21 for LXRα compared to −0.47 for PXR). We note that the highly unfavorable G10 preference for DR4 sites (Δz-score = −3.21) is not observed for DR1 sites (Δz-score = −0.65), demonstrating that this NR-specific preference is not shared across all spacer lengths (Supplementary Fig. 2.6). These results highlight the advantages of visualization of energy logos over traditional DNA binding logos, and demonstrate that novel base preferences can arise on DR sites of different lengths.

In NR-binding logos, we observe base preferences in the spacer sequence between DR half-sites (e.g., Figs. 2.2, 2.5f, positions 12–15). We note a strong preference for an adenine in the spacer sequence of DR1 sites, which has been demonstrated for PPAR and other NRs (Bolotin et al., 2009); however, such a distinct base preference is absent at longer spacer lengths (DR2–DR5). To investigate the contribution of the spacer sequence to NR specificity, we examined how spacer variants modulate NR-DNA binding (Fig. 2.5g). We focused our analyses on NRs that exhibit preferences for DR3 and DR4 sites. Examining the binding affinity distribution for SNVs within the spacer of a single seed sequence, we find that the spacer sequence can have considerable impact on binding affinity in an NR-specific manner (Fig. 2.5g), consistent with reports that NRs make DNA contacts with the spacer sequence (Lou et al., 2014). For instance, LXRα and LXRβ shows a drop in affinity (Δ z-score ~-2) for the DR4 spacer sequence 5'-GAGG-3', while this spacer sequence does strongly impact the affinity for THRα, VDR, or PXR

(Fig. 2.5g, top panel, red line). Given the established role for DNA shape in TF binding specificity (Yang et al., 2014; Zhou et al., 2015), we investigated whether DNA shape features in the spacer sequence might also contribute to the selectivity for different binding sites. We examined DNA shape features for spacer variants of DR3 and DR4 sites that enhance or diminish the binding of LXRα and VDR (Supplementary Fig. 2.7). The DNA shape features (i.e., major groove width, helix twist, propeller twist, and roll) examined are nearly identical for all comparisons. However, we observed a significant difference in the major groove width and roll parameters for VDR binding to DR3 sites. Our results suggest that DNA shape also plays a role in NR-binding specificity. Future studies that more exhaustively sample spacer sequences may enable identification of more subtle differences.

2.3.7 Genomic binding agrees with in vitro binding preferences of LXRα and PPARγ

Our NR-binding landscape (Fig. 2.2) shows DNA binding to DR sites with many spacer lengths. To determine whether NRs use these diverse sites in vivo, we evaluated the ability of our PBM-derived models to explain in vivo-bound regions from published ChIP-seq datasets (Methods). Examining published PPARγ binding data in HT29 colorectal cancer cells (GSE77039) (Savic et al., 2016), we find that all PPARγ models (DRs and half-sites) can discriminate bound regions from unbound. However, the DR1 model best describes the data (area under the curve (AUC) = 0.70, Fig. 2.6a), in agreement with established PPARγ binding preferences and our PBM data (Fig. 2.4). Testing other published DR1 models (Methods and Supplementary Data 3 from (Penvose

et al., 2019) (Isakova et al., 2017; Kulakovskiy et al., 2013; Matys et al., 2006), we find the HOCOMOCO-f1 DR1 model performs best (AUC = 0.67) and with similar accuracy to our DR1 model. These results suggest that binding to DR1 sites is an important determinant of in vivo PPARγ binding. In contrast, all models for LXRα yield similar AUCs (Fig. 2.6b), with the canonically preferred DR4 model performing similarly to the half-site models. Testing other published DR4 models we find JASPAR MA0494.1 (DR4) performs the best (AUC = 0.63), and performs similarly to PBM-derived half-site models (AUCs = 0.64). These in vivo binding results are consistent with our in vitro binding data, which show a strong DR1 preference for PPARγ and broader binding preferences for LXRα.

2.3.8 Functional sites agree with canonical spacer preferences for LXRα and PPARγ

We hypothesized that functional binding sites that regulate gene expression may have a different motif composition than the full set of genomic binding sites. Binding sites were annotated as 'functional' if they were located within 10 kb upstream of the transcription start site of genes whose expression changed >2-fold upon agonist treatment (GSE77039 (Savic et al., 2016), Methods). We then performed motif enrichment analysis for these functional PPARγ or LXRα binding sites. Strikingly, we observe an increase in the enrichment of the PPARγ DR1 and the LXRα DR4 models for their respective functional sites (Fig. 2.6c,d). These same trends are observed when we use alternate genomic constraints to define functional sites (i.e., 10 kb up- and downstream, or 50 kb upstream) (Supplementary Fig. 2.8). These results are consistent with a model wherein

NRs preferentially utilize DR full-sites at a canonical spacing for activating transcription, while genome-wide binding is determined by a broader set of DR and half-site sequences, consistent with our in vitro binding data.

2.3.9 LXRα binding via a half-site mode can drive gene expression

Our analyses reveal widespread binding of NRs to half-site sequences both in vitro and in vivo. Furthermore, I showed that half-site mode is utilized by NRs to bind not only to half-sites, but also to canonical DR sites. To determine whether NR half-site mode binding is functional and can drive gene expression, I assayed the ability of LXRα to activate a reporter gene from a binding site bound in a half-site mode on our PBM. Expression of luciferase reporter genes was monitored in HEK293T cells in the presence of over-expressed LXRα:RXRα and ligand or vehicle (Materials and Methods). I found that LXRα strongly induces gene expression, in a ligand-dependent manner, from a DR1 site (DR1.7) that is bound in a half-site mode by PBM (Fig. 2.7a, b, logo illustrates the 5'-half-site binding mode). Ablating the 5' half-site sequence (DR1.18) abrogates binding and drastically reduced reporter gene expression (from 36-fold induction down to 4-fold induction). Ablating the 3' half-site (DR1.17) does not affect binding affinity; however, unexpectedly, it strongly decreased reporter gene expression (from 36-fold induction down to 9-fold induction), demonstrating that in vitro affinity does not necessarily predict binding-site activity. Therefore, NRs binding via a half-site mode in vitro can drive gene expression, but DNA bases that do not affect binding affinity in vitro can affect function in vivo.

2.3.10 LXRα and PPARγ spacing preferences are defined by function not affinity

I next examined the ability of LXRα and PPARγ to promote gene expression from a range of DR1 and DR4 binding sites (Fig. 2.7c). In general, PPARγ drives higher levels of gene expression from DR1 sites, and LXRα functions better on DR4 sites, in agreement with their canonical spacer preferences. However, I see exceptions to these simple rules. First, LXRα can promote expression from the DR1.7 site (Fig. 2.7a) at a comparable or higher level than from the three DR4 sites (Fig. 2.7c). Second, for PPARγ, several high-affinity DR1 sites (DR1.8, DR1.3) show comparable or lower activity than the three DR4 sites, which are all bound with comparable or lower affinity. Complicating the interpretation, without NR overexpression, the DR4 sites exhibit lower reporter gene activity than DR1 sites (Supplementary Fig. 2.9). This low basal activity may exaggerate the NR-dependent activation determined for these sites, which is calculated as the fold-change between basal and NR-over-expressed conditions. Despite these complications, it is clear that affinity does not strongly predict activity of different NRs.

2.4 Discussion

Here we report the most comprehensive DNA binding dataset to date for the type II NRs, and provide a revised framework for interpreting NR-binding and regulatory specificity. I demonstrate more promiscuous DNA binding for NRs than has been previously reported, challenging the view that NR-binding specificity is defined solely by distinct DR spacer preferences. These findings agree with other PBM-based studies of NR homodimers that demonstrated nearly identical binding for RXRα and COUP-TF2,

and found that NR specificity does not solely depend on DR-spacing rules. We demonstrate that NR-binding-site activity does not follow binding affinity, and that the canonical NR DR spacer-length preferences better reflect activity rather than DNA-binding-site affinity. Our revised framework for NR-binding and function shows that NRs bind DNA via two binding modes to a broad set of DR and half-site sequences; this binding corresponds with in vivo binding, but does not correspond to in vivo function, which may involve additional layers of specificity (e.g., allostery) (Fig. 2.7d). Future studies that focus on refining the rules for NR-binding-site activity will clarify this general framework and improve genomic analyses aimed at predicting NR-dependent gene regulation, or the impact of SNPs on NR function, as in a recent analysis of PPARγ function.

Our study challenges the prevailing view that each NR heterodimer prefers binding to DR sites of specific spacer lengths. We show that all NRs can bind with high affinity to many DR spacer lengths in a full-site binding mode. Previous studies that sought to identify DR spacer preferences did not explicitly account for multiple NR-binding modes, potentially complicating their interpretations (Cotnoir-White et al., 2011; Evans and Mangelsdorf, 2014; Kurokawa et al., 1993; Perlmann et al., 1993). While we observe previously described DR spacer preferences, our study suggests a distinct biophysical interpretation for these preferences. We propose that DR preferences of NRs are not based on a large increase in binding affinity, but arise from a preference to bind in a full-site mode over a half-site mode, coupled with a moderate increase in affinity (i.e., LXRα and PPARα, Fig. 2.4). The implication that NR spacer preferences are primarily

about binding mode, rather than affinity, may provide a biophysical interpretation of NR preferences that links binding mode to in vivo function.

The disagreement between the promiscuous NR binding seen in our study and the canonical DR spacer preferences reported in the literature may be explained by differences in the approaches utilized. DR spacer preferences were initially characterized on a small number of DNA sequences obtained from promoter regions of genes that were upregulated upon ligand treatment, naturally biasing towards functional genomic binding sites (Cotnoir-White et al., 2011; Evans and Mangelsdorf, 2014; Kurokawa et al., 1993; Mader et al., 1993). Other high-throughput methodologies used to examine NR heterodimer binding preferences bias towards high-affinity binding sites and thus do not capture the full landscape of NR-binding specificity (Isakova et al., 2017). Our PBM approach, which queried the binding across a broad range of affinities and DR spacer lengths, reveals a more promiscuous NR-binding landscape.

Our NR-binding data are consistent with in vivo binding, and provide an updated framework for interpreting genome-wide binding data. For example, PPARγ ChIP-seq peaks are best modeled by a DR1 motif, consistent with the high-affinity binding observed for DR1 sites. In contrast, LXRα ChIP-seq peaks are modeled equally well by most DR models and half-sites (Fig. 2.6), consistent with broader in vitro specificity for LXRα. We note that a DR4 motif was identified by de novo motif analysis using this LXRα ChIP-seq dataset (Savic et al., 2016), but only when restricting the analysis to the highest scoring ChIP-seq peaks; when motif finding is performed on the full dataset, a half-site motif is identified. This example illustrates a source of confusion in the field:

reinforcement of established NR-binding preferences by conclusions supported by only a small fraction of the genome-wide binding data (Boergesen et al., 2012; Everett and Lazar, 2013; Savic et al., 2016). Re-interpreting the genomic data in light of our dataset, we find that the broader specificity found in vitro is consistent with in vivo binding.

Unexpectedly, we found that all type II NR heterodimers have the ability to bind DNA via a half-site mode on both full-sites and half-sites. This is a clear example of DNA-based allostery, in which interactions with DNA alter the structure of DNA-bound TFs. Allostery has been reported for the NRs (Gronemeyer and Bourguet, 2009; Meijsing et al., 2009; Schöne et al., 2016; Watson et al., 2013), and provides a mechanism to decouple affinity and activity. A provocative idea is that NR-binding mode can predict activity and explain NR functional preferences. Supporting this idea, a recent study of the glucocorticoid receptor (GR), a steroid hormone nuclear receptor, showed that GR homodimers can bind to half-site sequences in vivo to repress gene expression (Hudson et al., 2012). Our data on the preference of PPARγ and LXRα to bind in a full-site mode and drive gene expression from DR1 and DR4 sites, respectively, offer additional support for this idea. Other work has demonstrated that NR binding can be altered by coregulator proteins (Issa et al., 2001; Lefebvre et al., 1998), raising the possibility that NR binding modes may be altered in the presence of endogenous coregulators. Future studies that assess NR-DNA binding and binding modes in the presence of coregulators will help clarify the relationship between NR-binding mode, affinity, and activity. Our PBM dataset provides a valuable resource for these future studies aimed at elucidating the mechanisms of NR specificity in gene regulation.

2.5 Materials and Methods

2.5.1 Protein expression and purification

For all of the NRs used within this study, I cloned the human coding regions for the proteins into both N-terminal His-tag and GST-tag expression vectors for *E.coli* expression (pDEST 15 and pDEST17 vectors, respectively) and mammalian expression (pDEST26 and pDEST27, respectively). Over the course of three years, I made many attempts to purify all of the NRs myself, from both *E. coli* and from the human HEK 293 Freestyle cell line. I tested many different *E. coli* expression strains (BL21(DE3), C41(DE3), BL21(pRARE), and Arctic Express), tags (His, GST, FLAG) and expression conditions (temperature, density at induction, strength of induction). However, after three years I had only obtained full-length protein that showed sequence-specific DNA binding by EMSA for PPARγ and RXRα, and the decision was made to purchase the remaining NRs. Expression of full-length NRs in *E. coli* is known to be difficult as most of the protein ends up in inclusion bodies (Mossakowska, 1998). Expression in mammalian cells led to lower yield and purity, and we chose not to continue to pursue this method of expression. Throughout the expression and purification process, the yield and purity of the protein were checked by SDS-PAGE and western blotting (Supplementary Fig. 2.10 a-d). For samples for which the yield and purity was high, the activity of the protein (e.g, ability to bind DNA in a sequence-specific fashion) was tested using EMSA (Supplementary Fig. 2.11). For purchased proteins, the proteins were checked by western blot using an antibody against the epitope-tag (Supplementary Fig. 2.12) and with NR-specific antibodies before proceeding to PBM experiments. The type II NRs not included

in this study were left out because I was unable to purify or purchase full-length, active protein.

Full-length, wild-type human RXRα and PPARγ isoform 1 constructs were cloned into the Gateway vector pDEST17 (LifeTech) for propagation, mutagenesis, and expression. A TEV-protease recognition sequence was included between the coding sequence of the His-tag and RXRα and used to cleave the His-tag after purification. His-tagged RXRα and PPARγ were expressed using the BL21(DE3) *E. coli* strain (NEB). Transformed bacteria were propagated on Luria-Bertani broth (LB) plates supplemented with 100 μg/ml of carbenicillin. Protein expression was carried out in LB supplemented with 100 μg/ml of carbenicillin, with an initial outgrowth at 37 °C up to an OD of 0.4, transferred to ∼20 °C until they reached an OD of 0.6–0.7 and then induced with 1 mM IPTG. Protein was expressed at room temperature (∼20 °C) for 3 h. Cells were pelleted and stored at −80 °C until purification. Purification was carried out using HisTrapFF columns (GE Healthcare). The binding buffer was composed of 20 mM Tris HCl pH 7.4, 300 mM NaCl, 25 mM Imidazole, and 1 mM dithiothreitol (DTT) and the elution buffer was composed of 20 mM Tris HCl pH 7.4, 300 mM NaCl, 250 mM Imidazole, and 1 mM DTT. Buffers were supplemented with cOmplete Mini protease inhibitor tablets according to the manufacturer's instructions (Roche). Eluted fractions were analyzed by SDS-PAGE and fractions containing protein were combined. For PPARγ, the combined elution fractions were buffer exchanged into phosphate buffered saline pH 7.4 with 1 mM phenylmethylsulfonyl fluoride (PMSF) and 10% glycerol using an Amicon Ultra centrifugal filter (30k MWCO). Elution fractions of RXRα were dialyzed against three

changes of binding buffer. Next, the His-tag was cleaved from RXRα by overnight incubation at 4 °C with TEV protease (Sigma–Aldrich). After cleavage, the RXRα sample was re-purified as described above; however, this time the flow-through fraction from the column loading was collected and used in all PBM experiments, as this fraction contained the RXRα from which the His-tag was successfully cleaved. The combined flow-through fractions were buffer exchanged into phosphate buffered saline pH 7.4 with 1 mM PMSF and 10% glycerol using an Amicon Ultra centrifugal filter (30k MWCO).

The RXRα and PPARγ DNA binding domain mutants were made by site-directed mutagenesis using the NEB Q5 site-directed mutagenesis kit (New England Biolabs) following the manufacturer's instructions.

Primers used for the mutagenesis:

RXRα-DBDmut:

Forward = 5'-CTTCTTCTTCAAGGCGACGGTGCGCAAGGACCTG,

Reverse = 5'- CCCGCGCACCCCTCGCAGCTGTACACTCCATCAGC;

PPARγ-DBDmut:

Forward = 5'-CTTCCGGGCAACAATCAGATTGAAGCTTATCTATGACAG,

Reverse = 5'- AAACCCGCGCATCCTTCACAAGCATGAACTCCATAGTG.

For DNA binding domain mutant experiments, PPARγ, RXRα, PPARγ-DBDmut, and RXRα-DBDmut were expressed using the PURExpress IVT kit (NEB) according to manufacturer instructions. The concentration of all IVT-produced proteins was estimated

by western blot by comparison to purified proteins (Supplementary Fig. 10 d&e). All other purified proteins used were purchased (see Table 2.1 for details).

2.5.2 SDS-PAGE

For SDS-PAGE quality control experiments (Supplementary Figs. 2.10), proteins were separated by electrophoresis on 10-well 4-20% Mini-Protean TGX gels (Bio-Rad). Proteins were denatured in Laemmli sample buffer (Bio-rad) at 95 °C for 10 min. For raw lysate and elution fractions (Supplementary Fig 2.10a and 2.10b) 15 μl of lysate with 5 μl of 4x Laemmli. For the concentrated and buffer exchanged samples 0.5 μl of PPARγ was used (Supplementary Fig 2.10c) and 1 μg of RXRα was used (Supplementary Fig. 2.10d), each in a final volume of 20 μl. Precision Plus All Blue Pre-stained Protein Standard (Bio-Rad) was also loaded onto the gels. Gels were electrophoresed until the bromophenol blue dye front reached the bottom of the gel.

2.5.3 Western blotting

Proteins were detected by western blotting using antibodies that recognized either the protein or the epitope-tag. For western blot experiments, proteins were separated by SDS-PAGE electrophoresis and then transferred to 0.2 μm nitrocellulose membrane (Millipore) for 1.2 hr at 100 V. The membrane was rinsed with deionized water and then incubated in blocking buffer (3% milk in PBS pH7.4 with 0.05% Tween-20) for 1hr at room temperature. Membrane was rinsed 3 x 5 min in PBST (PBS pH 7.4 with 0.05% Tween-20). Next, membranes were incubated with primary antibody diluted in PBST for 2 hr at room temperature. Primary antibodies were used at a 1:2000 dilution and included

anti-RXRα (Active Motif 61029), anti-His (Sigma H1029), and anti-PPARγ (Abcam ab41928). After the 2 hr incubation in primary antibody, the membrane was rinsed 3 x 10 min in PBST and then incubated for 1 hr at room temp in horseradish peroxidase-conjugated goat anti-mouse antibody (Abcam ab205719) diluted at 1:5000 in PBST. Next, the membrane was rinsed 3 x 10 min in TBST. Membranes were developed in Supersignal west chemiluminescent substrate for 5 min (Thermo Scientific 34087). Chemiluminescence was detected on radiography film and developed on a Kodak M35 X-OMAT Processor (Kodak Diagnostic Imaging).

2.5.4 Radiolabeled Electrophoretic Mobility Shift Assay

Double-stranded (ds) DNA oligonucleotides (60 bp) were prepared by primer extension (45 µL total) using 8 µM of 60-bp ssDNA template (IDT Technologies, sequences given below), 8 µM of 24-bp extension primer , 1.6 mM dNTPs (New England BioLabs), and 1x ThermoPol Reaction Buffer (NEB). A separate enzyme mixture (5 µL total) was prepared with 4 units of Bst DNA polymerase- Large fragment (NEB) and 1x ThermoPol Reaction Buffer. Reaction mixtures were heated in a thermocycler to 95˚C, and gradually cooled to 63˚C by decreasing the temperature by -0.1˚C/s. The enzyme mix was placed in the thermocycler for 1 min to equilibrate to temperature, and then 5 µL of this mix was pipetted into each 45 µL reaction mixture. The reactions were incubated at 63˚C for an additional 90 min to allow completion of the primer extension reaction. The double-stranded DNA was purified from the extension reaction using a MinElute PCR Purification Kit (Qiagen) per the manufacturer's instructions. DNA probes were radioactively labeled using T4 polynucleotide kinase (PNK) (NEB) and ATP [γ32P]

(Perkin Elmer). For each probe, 2 pmol of oligonucleotide was incubated with 10 units of T4 PNK and 20 μCi ATP [γ32P] in 1x PNK Buffer (15 μL total reaction) @ 37˚C for 2 hr. After labeling, the probe was purified using a QIAquick Nucleotide Removal Kit (Qiagen) following the manufacturers guidelines to remove residual ATP [γ32P].

DNA Probes (5'→3'):

Note: For direct repeat sites, each half-site is bolded and underlined.

Extension primer: CCTTCATTCTACGCTGTCAATCGC

DR1:
GCCAAACT**AGGTCA**A**AGGTCA**GCCAACCAGCGATTGACAGCGTAGA
ATGAAGG

DR4:
GCCAAACT**AGGTCA**CGAA**AGGTCA**AAGTCGAATGACGCGATTGACA
GCGTAGAATGAAGG

Non-specific competitor:
GATAAGCGCCATTCAGGGGTCCACAGTTCACGTAGTGCGATTGACAGC
GTAGAATGAAGG

DNA-binding reactions contained 1 nM ^{32}P-labeled DNA in 1x EMSA Binding Buffer (1x PBS pH 7.4, 0.02% Triton-X-100, 1 mM dithiothreitol (DTT), 0.2 mg/mL bovine serum albumin (BSA) (NEB), and 5% v/v glycerol (Sigma). Protein concentrations used in each DNA binding reaction are listed in the figure (Supplementary Fig. 2.11). The DNA-binding reactions were incubated at room temperature for 30 min. Samples were electrophoresed for 1.7 hr at 70V on non-denaturing polyacrylamide gels (6% of 29:1 acrylamide:bisacrylamide [Fisher], 0.5x Tris Buffered EDTA [pH 8.3]. Gels were dried under suction for 1.5 hr at 80˚C, using a BioRad Gel Dryer. Dried gels were

placed onto a PhosphorImager (GE Healthcare) and left to expose overnight (~12 hours). DNA bands were visualized using a Typhoon Trio scanner (GE Healthcare) with a 100 μm pixel size.

2.5.5 PBM custom design

PBM experiments were performed using custom-designed microarrays (Agilent Technologies Inc. AMADID 084387, 4 × 180 K format). PBM probes contain a 24 nt constant primer region, a 34 nt variable region, and a 5' GC dinucleotide cap (probe sequences can be found in Supplementary Data 4 of (Penvose et al., 2019)). For each unique SNV probe sequence, five replicate probes were included in each orientation (10 probes per unique sequence). For all other probe sequences four replicate probes were included with the 34 nt variable region in each orientation (8 probes per unique sequence).

Seed probes: Direct repeat seed probes were generated from a variety of flank and half-site combinations. The starting half-site sequences were chosen from the literature as variations from the 5'-AGGTCA-3' consensus half-site (Literature curation of binding sites can be found in Table 2.2). Any given seed half-site varied from the consensus half-site by at most 2 bases. For SNV modeling we require that the seed sequence z-score > 3 and I chose not to vary the starting half-site more than 2 bases from the consensus half-site as more variation than this would likely lead to a low scoring seed sequence that would not be able to be modeled.

For each seed sequence, I created a corresponding spacer set at spacer lengths of 0-5 nt (DR0-5) (Seed-spacer sequence sets can be found in Table 2.6). For any given spacer set, the flanks and half-sites for that set are held constant and only the spacer length and sequence are varied. This allowed me to directly compare the affinity for different flank and direct repeat sequences at different spacer lengths. One limitation of this design is incomplete sampling of the spacer sequence at longer spacer lengths (DR3-DR5). Care was taken to design spacer sequences with as much variety as was possible given the space limitations on the PBM. Spacer sequences were taken from binding sites reported in the literature (Table 2.2) and directly designed to introduce starting sequence diversity.

SNV probes: DR seed sequences, defined by two 6-bp half-sites and a variable spacer (0–5 bp), were aligned within the 34 nucleotide variable region of each PBM probe. For each seed sequence, SNV probes were created that had a single-nucleotide variant at each position of the DR half-sites, the spacer sequence between the DR half-sites, and in the 5 bp flanks of each site. Therefore, for a single 13 bp DR1 site (i.e., $6 + 6 + 1 = 13$), including 5 bp flanks on either side, there would be 69 (i.e., 23×3) unique SNV probe sequences.

Half-site ablation probes: For each DR seed sequence, probe variants were created with each half-site ablated. Ablations were performed by identifying the position in the half-site that contributes most to the NR-binding score and replacing it with a penalizing base. Ablations probes were conceptually designed by Ashley Penvose and Jessica Keenan. Jessica Keenan did the scripting to create the probes.

Random genomic probes: 34 nt regions were randomly chosen from the UCSC hg19 build of human genome. Sequences were removed that contained Ns or single-nucleotide repeats longer than three nucleotides.

Other probes: This array design also included a variety of other probe types including DRs at longer spacer lengths (DR6-DR9), inverted repeats (IR0-9), and everted repeats (ER0-9). Due to space limitations on the 4x180k array, sites of these types were included as individual probes, rather than as full seed and SNV sets. The intention of this design was to gain insight into which alternative site arrangements show high affinity binding for each NR dimer and require further investigation in future studies. The exclusion of these probes from my analyses is discussed in section 2.6.1.

2.5.6 PBM experiments and analysis

See Table 2.1 for protein concentrations used for each NR. For all NR heterodimers, the RXR partner was initially tested at a concentration of 270nM. When experiments showed poor agreement between antibody replicates, which suggested we were measuring some non-heterodimeric binding, the concentration of the RXR partner was lowered. This was the case for PPARγ and THRα. For the PBM experiments on individual NRs, each NR was used at the same concentration as in the heterodimer experiment. RXRα alone was tested at 180 nM, as this would yield a homodimeric receptor concentration of 90 nM, which is the same concentration that was tested for the heterodimers.

Microarrays were double-stranded as previously described (PBM double-stranding primer 5'-CCTTCATTCTACGCTGTCAATCGC-3') (Berger and Bulyk, 2009; Berger et al., 2006a). All washes were performed in coplin jars on an orbital shaker at 125 rpm. Double-stranded microarrays were first pre-wetted in phosphate buffered saline (PBS) containing 0.01% Triton X-100 for 5 min, rinsed in a PBS bath, and then blocked with 2% milk in PBS for 1 h. After blocking, arrays were washed in PBS containing 0.1% Tween-20 for 5 min, then in PBS containing 0.01% Triton X-100 for 2 min and then rinsed in a PBS bath. Proteins were then incubated on the array for 1 h in a binding reaction containing: PBS pH 7.4 with 2% milk, 0.02% Triton X-100, 1 mM DTT, 0.2 mg/ml bovine serum albumin, and 0.4 mg/ml salmon testes DNA (Sigma D7656). Preliminary PBM experiments for PPARγ:RXRα and RXRα were performed with and without the ligands rosiglitazone and 9-*cis* retinoic acid, respectively, and we found no change in NR binding; therefore, all experiments were performed in the absence of ligand. Following the protein incubation, microarrays were washed with PBS containing 0.5% Tween-20 for 3 min, then in PBS containing 0.01% Triton X-100 for 2 min followed by a brief PBS rinse. Microarrays were then incubated with 20 μg/ml of primary antibody in 180 μL of 2% milk in PBS for 20 min. For heterodimers, separate experiments were performed using an antibody against each protein within the heterodimer. In all experiments, anti-RXRα antibody (Active Motif 61029) was used to detect RXRα and anti-His antibody (Sigma H1029) was used to detect the NR partner with the following exceptions: anti-PPARγ antibody (Abcam 41928) was used in all experiments with PPARγ, and Alexa488-conjugated anti-GST antibody (Life Tech

A11131) was used for all PPARα experiments. Excess primary antibody was removed by washing with PBS containing 0.05% Tween-20 for 3 min and then in PBS containing 0.01% Triton X-100 for 2 min. Arrays were next incubated with 20 μg/ml of Alexa488-conjugated secondary antibody (anti-mouse A488, Life Tech A11001) in 180 μL of 2% milk in PBS for 20 min (PPARα was probed with an Alexa488-conjugated anti-GST primary antibody as described above and did not require a secondary antibody). Excess antibody was removed by washing 2x with PBS containing 0.05% Tween-20 for 3 min and then in PBS for 2 min. Microarrays were scanned with a GenePix 4400 A scanner and fluorescence was quantified using GenePix Pro 7.2. Exported data were normalized using MicroArray LINEar Regression (Berger et al., 2006a). Microarray probe sequences and fluorescence values from each experiment are provided (See Supplementary Data 4 of (Penvose et al., 2019)). NR dimers exhibit an orientation-specific bias in our PBM experiments; therefore, data from probes in a single orientation (i.e., '_o1′ probes in Supplementary Data 4 of (Penvose et al., 2019)) was used in our final analysis. However, all results were observed for probes in both orientations and models from each orientation showed good agreement.

Position frequency matrices (PFMs) and DNA-binding logos were generated for each seed sequence with z-score >3.0 using the previously described SNV-based approach (Andrilenas et al., 2018), with β set to 15/maximum z-score. Briefly, logos for single seed sequences are generated using the binding data to each seed sequence and all the single-nucleotide variant (SNV) sequences for that seed sequence. For a binding site of length L there will be 3xL SNV sequences. Logos for an NR binding to a specific DR

spacer length are determined by averaging over the individual seed sequence logos. To generate logos for a specific DR spacer length (Fig. 2.2), PFMs for all seed sequences at that spacer length were clustered into full-site, 5'-half-site or 3'-half-site PFMs. Average PFMs of each type (i.e., full, 5'-half-site or 3'-half-site) were then generated by directly averaging over the individual PFMs (i.e., averaging individual matrix elements and normalizing each column to 1). As the half-site PFMs are the same length regardless of the starting DR seed length, the final 5'-half-site and 3'-half-site PFMs were further averaged over PFMs generated at all spacer lengths. The z-score energy matrix (Fig. 2.5f) was generated in the same manner, without the initial transformation from z-score to frequency (Andrilenas et al., 2018).

2.5.7 Reporter gene assays

PPARγ, LXRα, and RXRα were cloned into the N-terminal His-tagged protein mammalian expression plasmid pDEST26 (LifeTech). Reporter constructs for test sequences were ordered synthesized (Twist Bioscience) and were flanked by two BsaI cut sites, which were used to clone the sequences into pNL3.1-minP/Nluc (Promega). All sequences tested can be found in Table 2.5. HEK293T (ATCC) cells were cultured in DMEM (Gibco 11965-092) supplemented with 10% FBS (Gibco 26140079). Cells were plated in tissue culture treated 96-well plates seeded at a density of 12,500 cells per well and allowed to adhere overnight. PEI:DNA complexation reactions were prepared at a ratio of 3:1 (PEI:DNA) in 500 µl of Opti-MEM (Gibco 51985-034) and allowed to complex for 20 min at room temperature. Each 96-well plate well received 20 µl of transfection mixture containing 16 ng of total plasmid: 1 ng of transfection normalization

plasmid (pGL4.54-Luc2/TK); 10 ng of reporter plasmid (pNL3.1-minP/Nluc); and either 5 ng of empty pDEST26 for the no overexpression conditions (NoOE); or 2.5 ng of RXRα in pDEST26 combined with either 2.5 ng of PPARγ or LXRα in pDEST26 for protein overexpression condition (OE). Twenty-four hours after transfection, 80 µl of media was removed from each well and replaced with 80 µl of fresh media containing the appropriate ligand treatment. PPARγ ligands were 1 µM rosiglitazone (Sigma–Aldrich) and 1 µM T0070907 (Sigma–Aldrich). LXRα ligands used were 1 µM GSK2033 (Sigma–Aldrich) and 500 nM T0901317 (Sigma–Aldrich). Luciferase activity was assessed 18 h after addition of the ligand using the Nano-Glo Dual Luciferase reporter assay system (Promega). Dual luciferase signal was quantified using a VICTOR-3 plate reader (PerkinElmer). To control for transfection efficiency, the Nanoluc (Nluc) reporter plasmid signal was normalized to the constitutive luciferase signal (Luc2) (i.e., signal from pGL4.54 plasmid) (Nluc/Luc2). Normalized signal for all test DNA elements were then further normalized to empty vector (pNL3.1-Nluc with an insert of equal length to test sequences but lacking any half-site or direct repeat sequences). Fold-induction values for each protein + reporter combination were calculated relative to the background activity of each reporter plasmid in the absence of protein overexpression: (protein + reporter)/(control + reporter) = OE/NoOE (Supplementary Fig. 2.9). Reporter assays were performed as three biological replicates with three technical replicates per biological replicate.

2.5.8 Competition Electrophoretic Mobility Shift Assay (EMSA) experiments

Complementary DNA oligonucleotides (from Integrated DNA Technologies, sequences in Table 2.4) were annealed in a thermocycler by raising the temperature to 98 °C and reducing the temperature by 0.1 °C/sec until a temperature of 4 °C was reached. EMSA buffer formulation for all reactions was 1x PBS pH 7.4 with 0.2% BSA, 5 mM DTT, 10% glycerol, and 0.02% Triton-X100. For the direct binding experiment, 1 nM of IR700-labeled P1 probe (Integrated DNA Technologies) was incubated with varying concentrations of PPARγ:RXRα in a 20 µL reaction. For competition experiments, 2 nM of IR700-labeled P1 probe was incubated with PPARγ:RXRα (12 nM:4 nM) in a 20 µL reaction with various concentrations of unlabeled competitor sequences (0, 0.2, 0.63, 2, 6.3, 20, 63, 200, 630, and 2000 nM). Reactions were incubated for 1 h at room temperature and then electrophoresed in 0.5x Tris Boric Acid EDTA Buffer (TBE) on a 6% TBE-acrylamide gel at 50 V for 3 h. Gels were scanned on the Odyssey CL-X (LI-COR) at 84 µm resolution. Fluorescence of the shifted band was quantified using ImageStudioLite software. All K_d calculations were done with DynaFit 4 software (Kuzmič, 1996) using a previously described competition protocol (Golden et al., 2013). Percent competition was calculated by the formula:

$$\% \text{ inhibition} = (F_0 - F_c)/F_0 * 100$$

F_0: fluorescence of shifted band with no competitor DNA

F_c: fluorescence of shifted band at given concentration of competitor DNA.

2.5.9 Enrichment of NR-binding sites in ChIP-seq data

Receiver-operating characteristic (ROC) curve analyses were performed to quantify the extent to which NR-bound (true positive) regions scored more highly than unbound (true negative) regions with PWM models. True-positive regions for LXRα and PPARγ were derived from ChIP-seq data from HT29 colorectal cancer cells (GSE77039) (Savic et al., 2016). ChIP-seq was available for two biological replicates of HT29 cells treated with agonist (GW3965 for LXRα or rosiglitazone for PPARγ) for 2 h and 48 h. For each NR, ChIP peaks with 50% reciprocal overlap within time points and between time points were considered true-positive regions. True-negative regions were derived from DNase-seq of HT29 cells (GSE90403) (Consortium, 2012). Regions with 50% reciprocal overlap between the two available DNase-seq biological replicates were identified, and all ChIP peaks from the corresponding NR ChIP datasets were then subtracted from the DNase-seq regions. Regions matched in size to each ChIP-derived true-positive region were randomly chosen from ChIP-subtracted DNase-seq regions to create the true negative regions. Background nucleotide frequencies for calculating PWMs from PFMs were taken from the nucleotide distribution of the DNase-seq regions with 50% reciprocal overlap between the two replicates. To score sequences, the following formalism was used:

$$p_{i,j=}\frac{f_{i,j}+sb_i}{\sum_i f_{i,j}+s}$$

Probability of an A,C,G or T ($i = 0,1,2,3$ respectively) occurring at position j of the sequence being evaluated.

$f_{i,j}$: frequency defining the position frequency matrix

b_i: nucleotide background frequencies: A: 0.24; T: 0.24; C: 0.26; G: 0.26

s: pseudo-count to deal with zeros (s = 0.001)

The PWM score is the sum over all base positions (j) of the corresponding $S_{i,j}$ values for a particular sequence:

$$s_{i,j} = log_2(\frac{p_{i,j}}{b_i})$$

Area under the ROC curve (AUC) values are reported to quantify the enrichment, and a Wilcox-Mann-Whitney (WMW) U test was applied to calculate the significance of each AUC value. AUC and WMW U test values were calculated in the R statistical package using the wilcox.test function. All manipulations of genomic regions (identification of overlapping regions, region subtractions, etc.) were performed with BEDTools 2.26.0 (Quinlan and Hall, 2010).

To examine the motif enrichment of currently available models, we performed the ROC analyses described above with publicly available PFMs. Each PFM was normalized such that the nucleotide frequencies at each position sum to 1. The following models were used: LXRα (MA0494.1(Khan et al.); HOCOMOCO f1 (Kulakovskiy et al., 2013)),

PPARγ (Isakova et al., 2017); M00512, M00515, M00528 (Matys et al., 2006); MA0065.1, MA0065.2, MA0066.1 (Khan et al.); HOCOMOCO f1, HOCOMOCO s1 (Kulakovskiy et al., 2013)).

To examine motif enrichment for putative 'active' sites near differentially expressed genes, RNA-seq data from HT29 cells (Savic et al., 2016) were used to identify regions that are likely to be actively controlling transcription. We re-analyzed the published RNA-seq data using DESeq2 (Love et al., 2014) to identify genes upregulated upon agonist treatment compared to vehicle only (DMSO). Transcripts with a fold-change greater than 2 and adjusted *p*-values less than 0.01 were considered upregulated. For PPARγ, transcripts upregulated after both 24 and 48 h of rosiglitazone treatment were considered for further analysis. For LXRα, transcripts upregulated after 48 h of GW3965 and T0901317 treatment were considered for further analysis. For each NR, ChIP regions with 50% reciprocal overlap between replicates and time points and within the indicated regions associated with upregulated genes were considered active true positives for enrichment analysis. Regions matched in size to each active region were randomly chosen from the true-negative regions described above to create the true negative regions. ROC analyses were performed as described above.

2.5.10 DNA shape analysis

Binding to spacer-sequence variants of five DR3 and five DR4 seed sequences was analyzed (Supplementary Fig. 2.7). For each DR3 seed sequence, the PBM z-scores of the seed sequence and corresponding 9 SNV sequences (i.e., sequences with base variants at positions B1, B2, or B3) were analyzed to identify the two highest affinity and

the two lowest affinity sites for each of the five seeds, resulting in a total of ten high and ten low-affinity spacer variants. The same procedure was performed for the DR4 sequences and the corresponding 12 SNVs at positions B1, B2, B3, and B4. For each of the 10 spacer variants, the following DNA shape parameters were calculated at each base position using the TFBSshape server (Yang et al., 2014): major groove width (MWG), helix twist (HelT), propeller twist (ProT), and roll. The distribution of the DNA shape parameters associated with high and low-affinity sequences were compared at each base position using a two-tailed *t*-test.

2.6 Rationale and Limitations

2.6.1 Rationale and limitations of our custom NR PBM design

The PBM approach for measuring the DNA binding affinity of TFs has many advantages including the ability to capture TF-DNA interactions over broad range of affinities (Andrilenas et al., 2015). Our custom-designed NR PBM allowed us to directly compare the binding of the NRs to the same set of sequences without introducing the amplification biases that are common with many other high-throughput methodologies to measure protein-DNA interactions (Andrilenas et al., 2015).

However, as is true with any technique, there are also a variety of limitations presented by the PBM methodology. As PBMs have a limited number of probes (~180k for a 4-chambered array), the first challenge in creating a custom-designed PBM is deciding what types of sequences to include on the array. We ultimately chose to use a seed and SNV design as this would allow us to generate models for individual sequences.

This SNV approach to modeling yielded the unique insight into the ability of NRs to utilize multiple binding modes to engage with DNA. However, as each seed and SNV set takes up a significant amount of space on the array (length of probe to be included in SNVs x 3 base variants in each position + 1 probe for starting seed), I was constrained in the number of seeds I could include at each spacer length. For longer spacer lengths (DR3-5), this space limitation on the PBM meant that I could not exhaustively test all possible sequences in the spacer. The sequence of the spacer is known to impact the DNA binding affinity of NRs, as NRs make sequence-specific contacts with this region of DNA (Bolotin et al., 2009, Lou et al., 2014). It is possible that some NRs could have shown high-affinity binding to other spacer lengths if different sequences had been used in the starting sequence of the spacers.

Due to space limitations, I was also unable to include seed and SNV sets for other half-site arrangements (IRs, EREs, and DRs at longer spacer lengths), all of which have been shown to be utilized by NRs. My custom-designed PBM contained many IR, ERE and DR6-9 probes, which were intended to be used to clarify which NRs could utilize these alternative site structures. However, on our DR seed and SNV probes we unexpectedly found that high-affinity binding occurs via a half-site mode even on sites designed as DRs. This meant that for sites designed as IRs, EREs, and longer DRs the binding we measured may have been occurring via a half-site mode and might not require the full IR, ERE, or DR-sites. For this reason, analysis of these probes was excluded from my thesis. Further studies utilizing a different custom-designed PBM would be necessary to determine the importance of these binding sites.

While space was a limiting factor on this custom-designed PBM, from the outset of this project we intended to generate PWM models of binding affinity as they remain the prevailing model used to represent protein-DNA interactions. PWMs assume statistical independence between positions across a sequence and as such exhaustive combinatorial sampling of all possible sequences is not necessary to generate these types of models.

2.6.2 Rationale and limitations of reporter assay experimental design

Eukaryotic gene transcription is complex process that is dependent on TFs recognizing specific DNA sequences across the genome (Smale, 2001; Evans and Mangelsdorf, 2014). An outstanding problem in understanding TF biology and gene regulation is to understand the relationship between TF binding and TF function (Evans and Mangelsdorf, 2014). The simplest model to explain TF function at any given TF binding site would be a direct correlation between the affinity of a TF for a given DNA binding site and the amount of transcript produced from that binding site. To test the hypothesis that transcriptional output correlates with TF binding-site strength, we chose to use luciferase reporter assays using a vector containing a 3x repeat for each given test DNA sequence with 20 bp of intervening sequence between each repeat. The sequence context of these binding-sites was held constant (i.e, the spacing sequences used between the 3x repeats were the same for all test sequences). This configuration does not mimic a genomic environment; however, it is the clearest way to explicitly test for a correlation between NR binding-site strength and transcriptional activity, as it allows us to control for other variables that are known to influence gene regulation such as cooperative

binding with other TFs and chromatin context (Everett and Lazar, 2013). In the context of this experimental setup, a lack of correlation between binding-site strength and function suggests that higher-order models incorporating other parameters (e.g. ability to recruit coregulators) are required to model eukaryotic TF function on individual DNA binding sites, and suggests that binding site strength may not be an important parameter in models of TF function, other than perhaps to predict occupancy of a binding site at any given concentration of a TF.

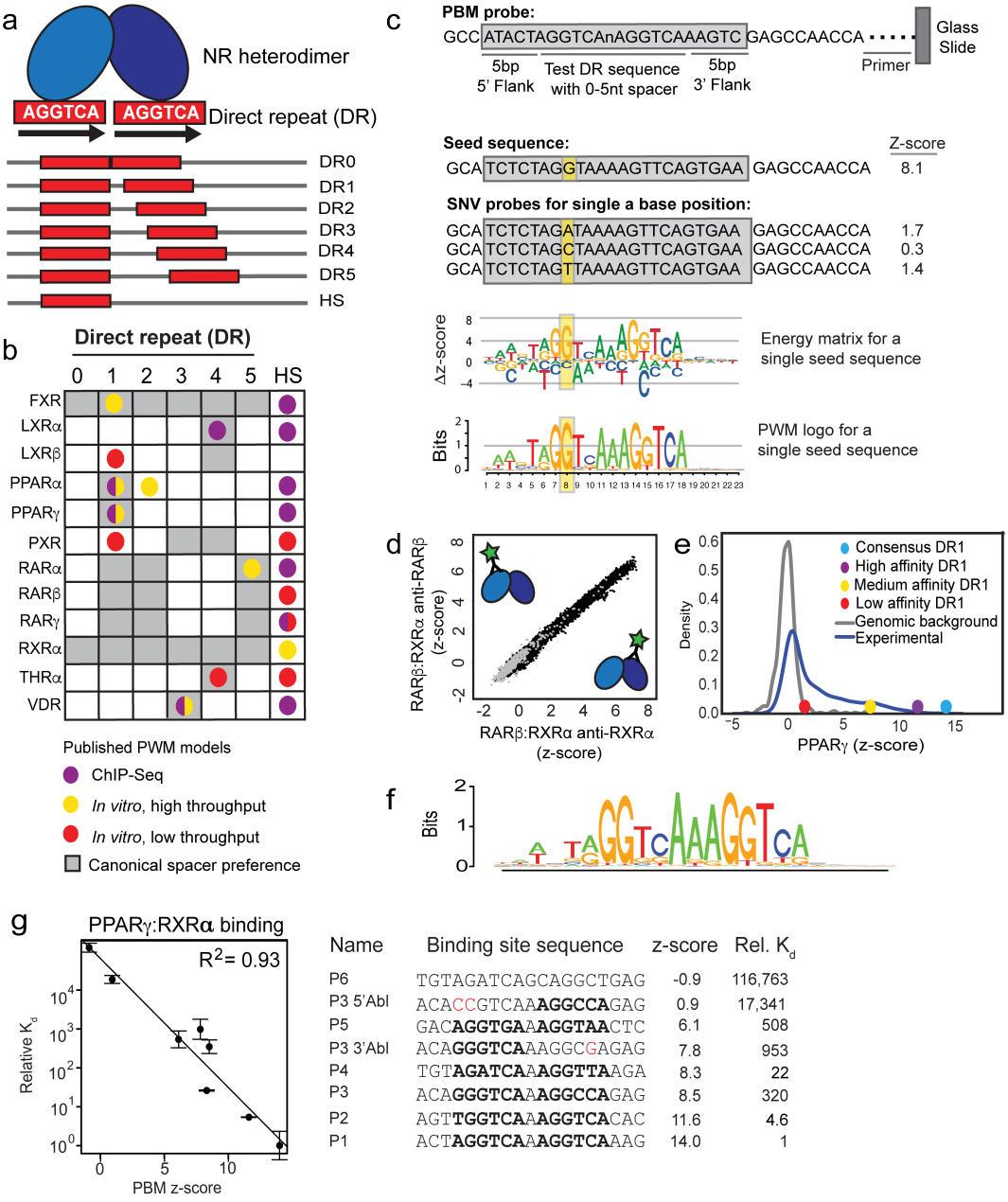


Figure 2.1: Characterizing NR-DNA binding with PBMs.

a Schematic of spacer preferences for NRs to direct repeats (DRs) and half-sites (HS). **b** Canonical spacer preferences of NRs indicate preferred spacer lengths from the literature (Table 2.2). Published PWM models are shown in colored dots that indicate the methodology used to derive the model (List of models can be found in Supplementary Data 3 of (Penvose et al., 2019). **c** Schematic of PBM probes, SNV probe organization and SNV-based motif generation for a single seed sequence. **d** Scatter plot of z-scores for RARβ:RXRα experiments detected with antibodies against each heterodimer partner. Dots represent average over ~5 replicates for all 10,728 unique SNV probes (black dots) and 500 background probes (gray dots) **e** PBM replicate

averaged z-score distributions for PPARγ:RXRα to all SNV probes. Z-scores for consensus DR1 and reported functional binding sites are highlighted (Table 2.3) (Juge-Aubry et al., 1997). **f** DR1 DNA-binding logo for PPARγ:RXRα generated from all DR1 full-site models from the PBM experiments. **g** Comparison of PPARγ:RXRα PBM z-scores and competition EMSA-determined relative K_d measurements for binding sites spanning a wide affinity range. Relative K_d values are normalized to the highest affinity sequence (P1) and represent mean over two independent experiments (error bars = STDEV). Identifiable DR half-sites in each binding sequence are shown in bold. Mutations introduced to ablate the 5' half-site of P3 (P3 5'Abl) or the 3' half-site of P3 (P3 3'Abl) are shown in red. Data and analysis by Ashley Penvose.

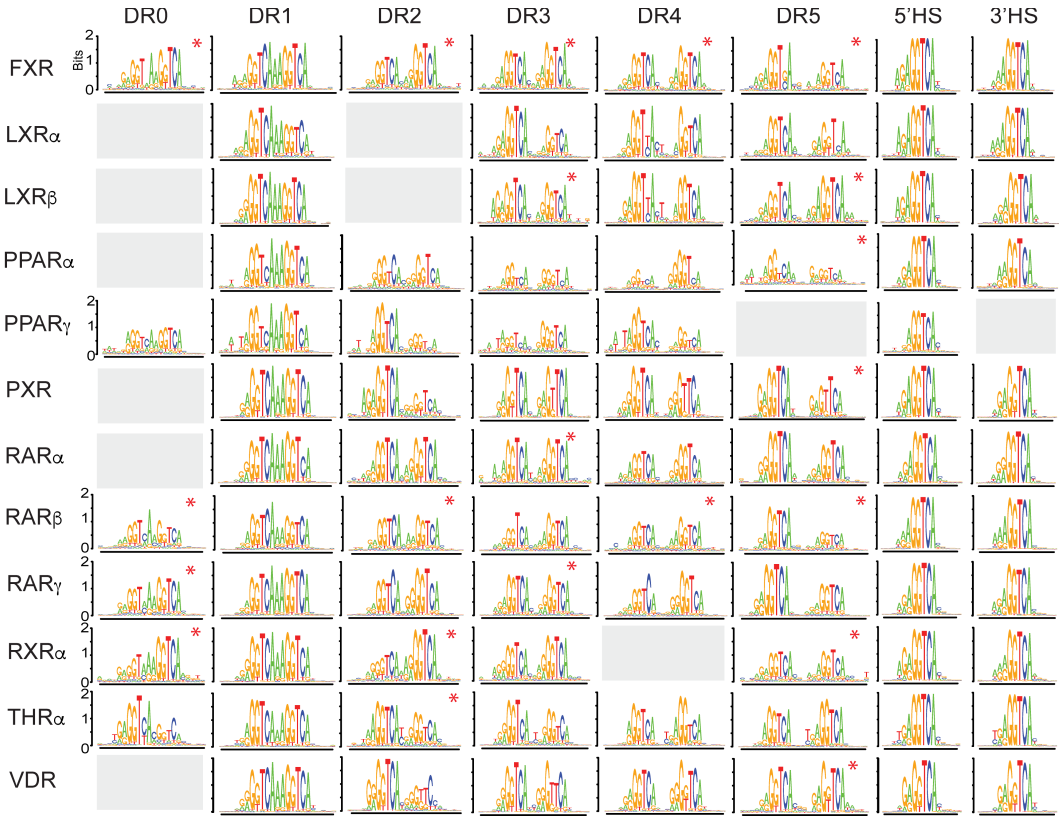


Figure 2.2: NR-binding specificity and DR preferences.

PBM-derived DNA-binding logos for 12 NRs at all examined DR spacer lengths. To generate a logo, the seed sequence must have a z-score > 3.0; grey boxes indicate that none of the seed sequences met this threshold at that spacer length for that NR. For each NR at each spacer length or half-site, we show the average logo (average of individual logos generated from each seed sequence with z-score > 3.0 at that spacer length). Logos based upon a single significant (z-score > 3.0) seed sequence are indicated (∗). Data and analysis by Ashley Penvose.

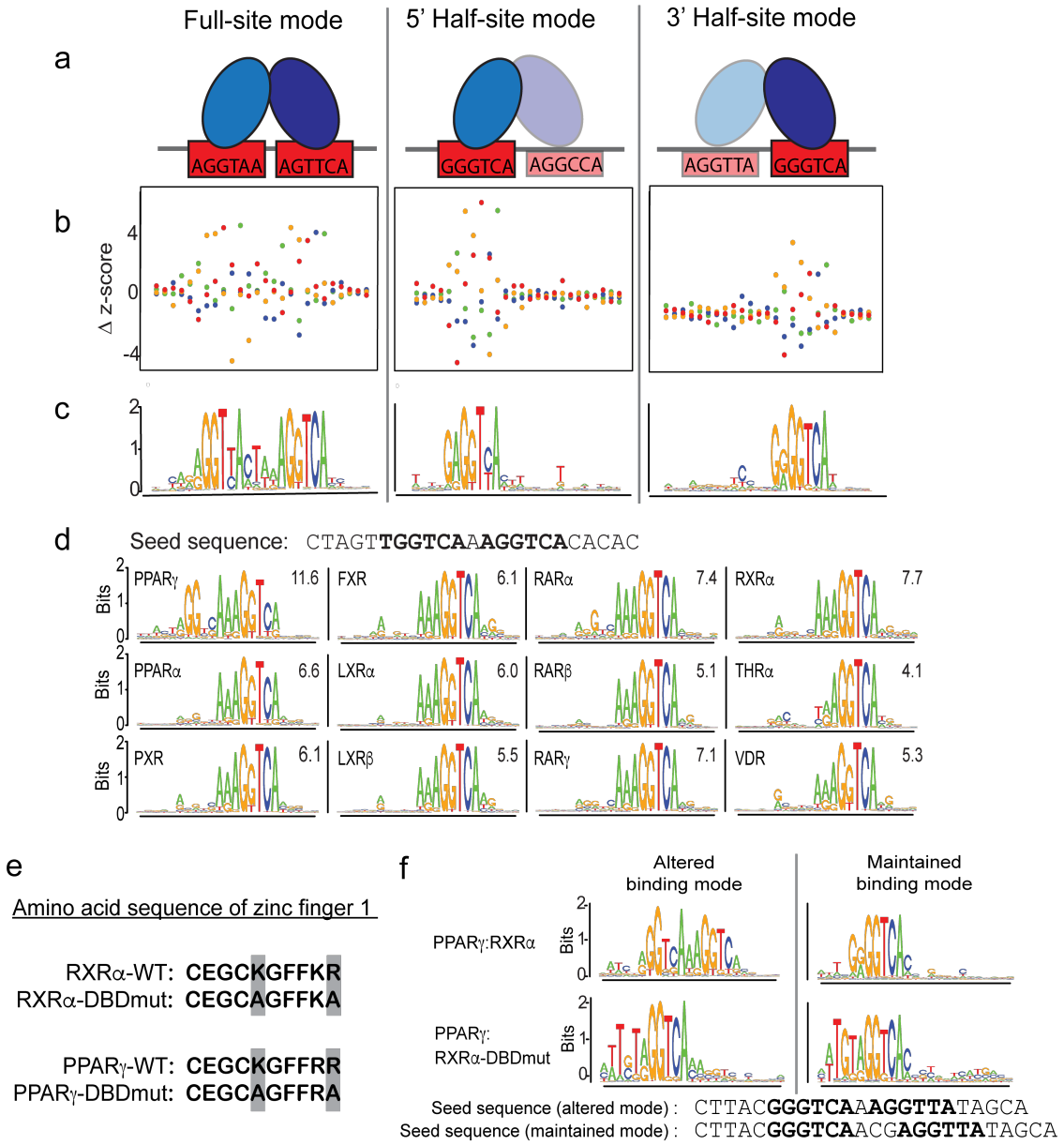


Figure 2.3: NR half-site binding mode.

a Schematic of NR full-site or half-site binding modes used by NRs to engage with DNA on the PBM. Faded protein and sequence indicates that the NR is not making base-specific contacts on the indicated half-site. **b**, **c** For three seed sequences bound with different modes, the impact of SNVs on LXRα heterodimer binding and the corresponding DNA-binding logos are shown. Binding perturbation for each SNV is shown as a Δz-score from the median z-score of all four base variants at each position. Colors correspond to base identity indicated in logos below. **d** DNA-binding logos for all 12 NRs generated for the single DR1 seed sequence shown. **e** Amino acid sequence of zinc finger 1 for the wild-type RXRα, RXRα DNA-binding domain mutant, wild-type PPARγ, and the PPARγ DNA-binding domain mutant. Altered amino acids are highlighted in

gray. **f** DNA-binding logos for individual seed sequences (shown) for which the binding mode was either altered (left) or maintained (right) for the PPARγ:RXRα-DNA binding domain mutant. Data and analysis by Ashley Penvose.

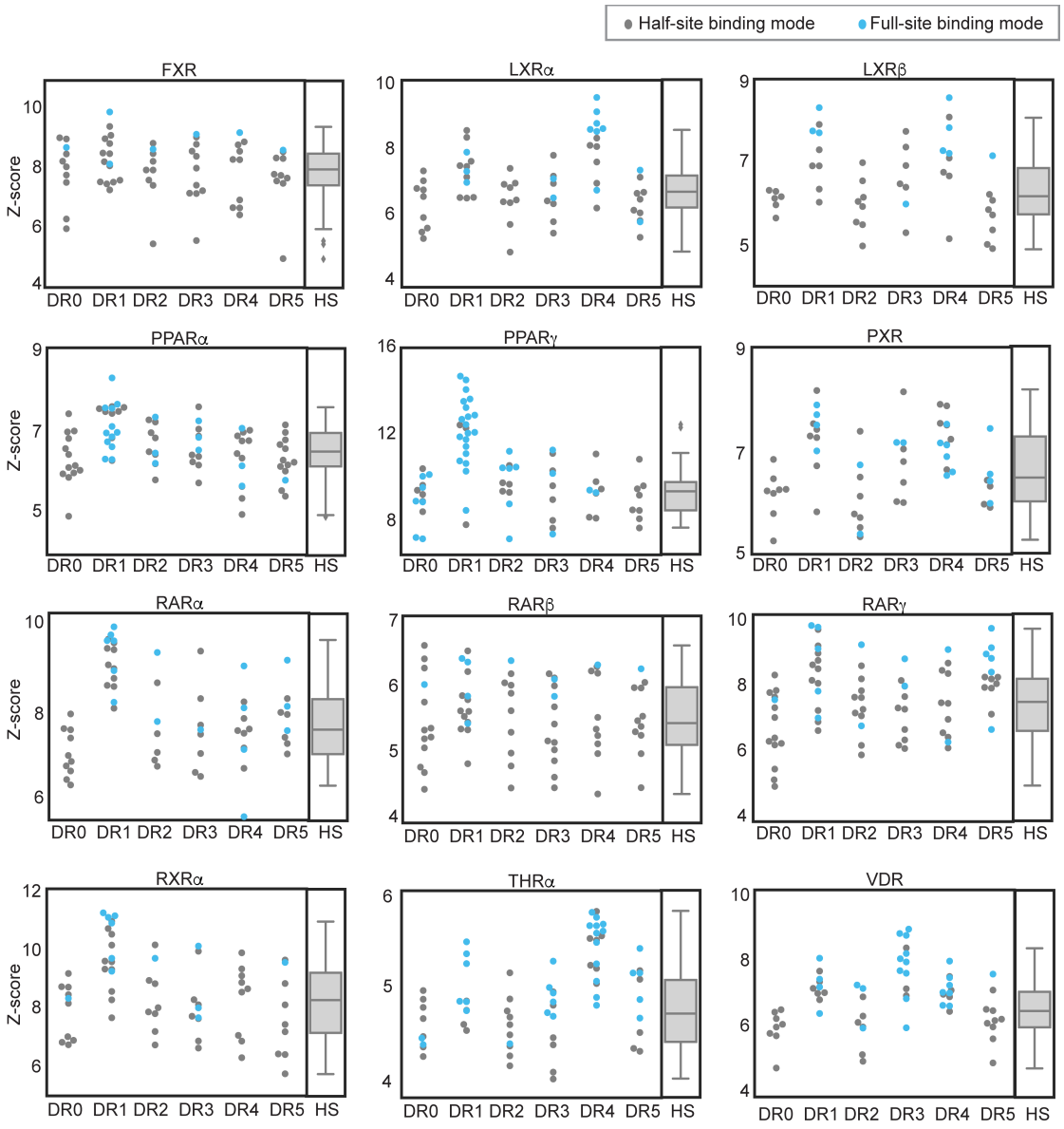


Figure 2.4: NR-binding affinity and mode for sequences at each DR spacer length.

At each spacer length, the replicate averaged z-score of the highest scoring SNV for each seed sequence is shown; seed sequences with z-score < 3 are not represented. Colors indicate binding mode for each seed sequence. For each NR, box plots show the z-score distributions for all sites that are bound in half-site modes across all direct repeat spacer lengths (the aggregate of all gray dots). Center line: median; box limits: upper and lower quartiles; whiskers: last datum within 1.5x interquartile range. Data and analysis by Ashley Penvose.

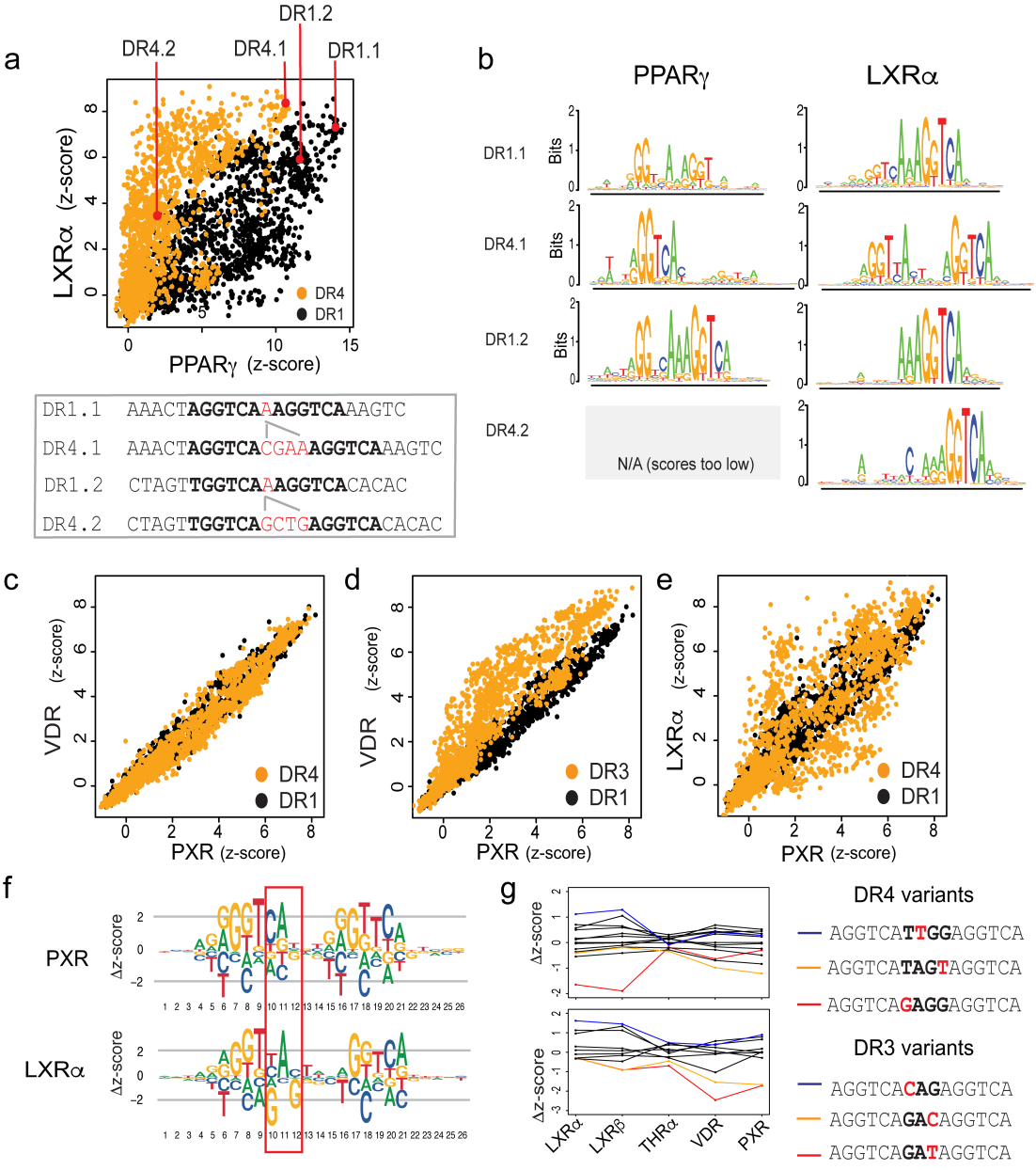


Figure 2.5: NR specificity differences.

a Scatter plots of LXRα and PPARγ binding to DR1 and DR4 sites. Each spot is the average of ~5 replicates for each unique DNA sequence (~1600 at each spacer length) on the PBM. DR1 and DR4 spacer-variant sequences are shown in the box below panel. **b** Binding logos generated for LXRα and PPARγ for the spacer-variant seed sequences from **a** are shown. **c** Scatter plots as in 2.**5a** of VDR and PXR binding to DR1 and DR4 sites. **d** Scatter plots as in 2.**5a** of VDR and PXR binding to DR1 and DR3 sites. **e** Scatter plots as in 2.**5a** of LXRα and PXR binding to DR1 and DR4 sites. **f** DR4 z-score logos, directly representing Δz-scores of SNV binding, are shown

for LXRα and PXR. Δz-scores are calculated (separately for each NR) as the difference from the median of all SNV variants. **g** Differential binding of NRs to spacer-sequence variants. (Top panel) Binding is shown for five NRs to the DR4 seed sequence 5'-AGGTCATAGGAGGTCA-3' and all 12 SNVs of the spacer region (spacer region in bold). Δ z-scores are calculated as in **2**.**5f**. (Bottom panel) Binding is shown for five NRs to the DR3 sequence 5'-AGGTCAGAGAGGTCA-3' and all nine corresponding SNVs of the spacer region (spacer region in bold). Examples of highly variant spacer sequences are indicated. Data generated by Ashley Penvose with this analysis performed by Trevor Siggers.

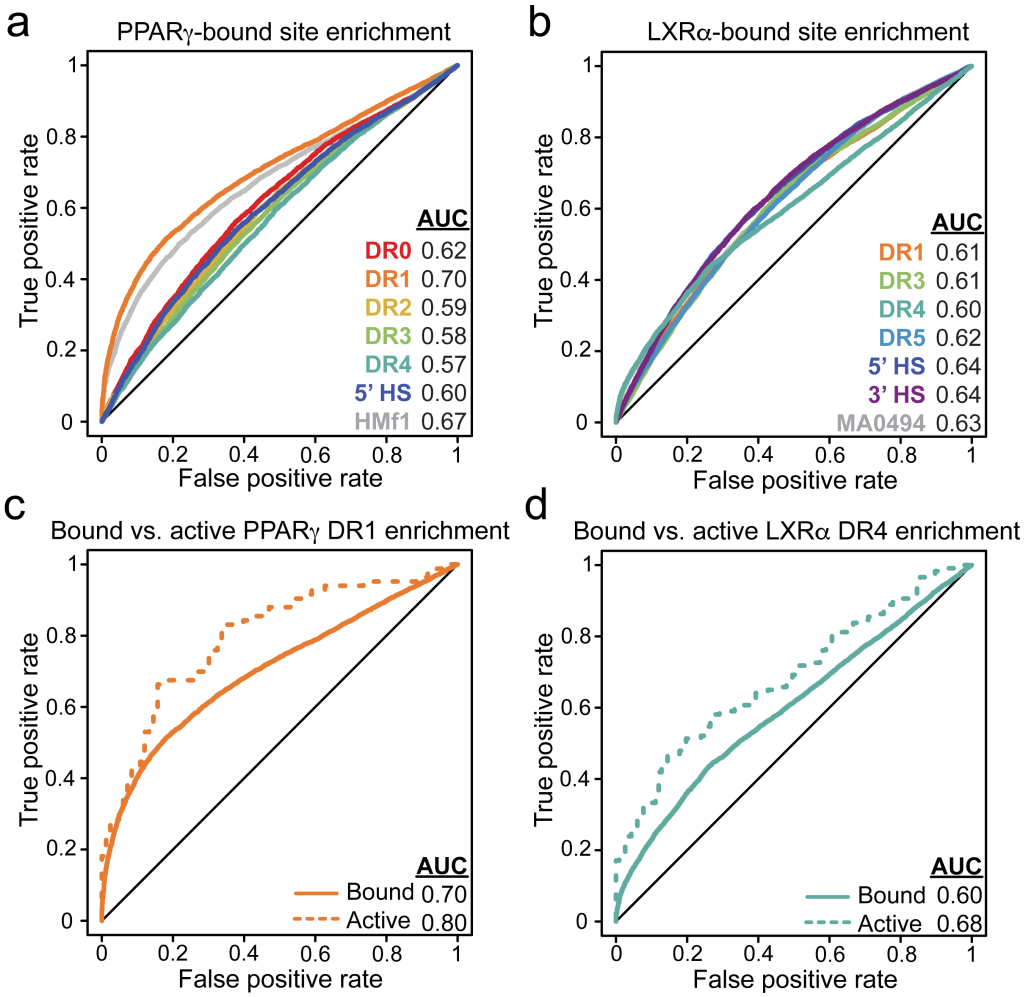


Figure 2.6: Genomic enrichment of NR-binding motifs.

a Receiver-operating characteristic (ROC) curves for PPARγ motif enrichment in ChIP-seq data. ROC curves and area under the curve (AUC) values for different PBM-derived NR-binding models are shown, along with the results for best-performing published PPARγ DR1 motif (HOCOMOCO-f1, HMf1 (Kulakovskiy et al., 2013)). Motif enrichment for all models had *p*-values < 10^{-46}, using a Wilcoxon rank sum test with continuity correction and Bonferroni corrected for multiple hypotheses. **b** ROC curves for LXRα motif enrichment in ChIP-seq data. ROC curves and AUC values for different LXRα binding models are shown. Results for best-performing published LXRα DR4 motif (JASPAR MA0494.1 (Khan et al., 2018)) are shown. **c** ROC curves and AUC values are shown for PPARγ DR1 motif enrichment in reproducibly-bound PPARγ ChIP-seq peaks (solid lines, Methods), and for those peaks occurring within 10 kb upstream of differentially expressed genes (i.e., active peaks). **d** ROC curves and AUC values are shown for LXRα DR4 motif enrichment in reproducibly-bound LXRα ChIP-seq peaks (solid lines, Methods), and for those peaks occurring within 10 kb upstream of differentially expressed genes (i.e., active peaks). The data/models tested in this analysis were generated by Ashley Penvose. The genomic enrichment analysis was performed by Jessica Keenan.

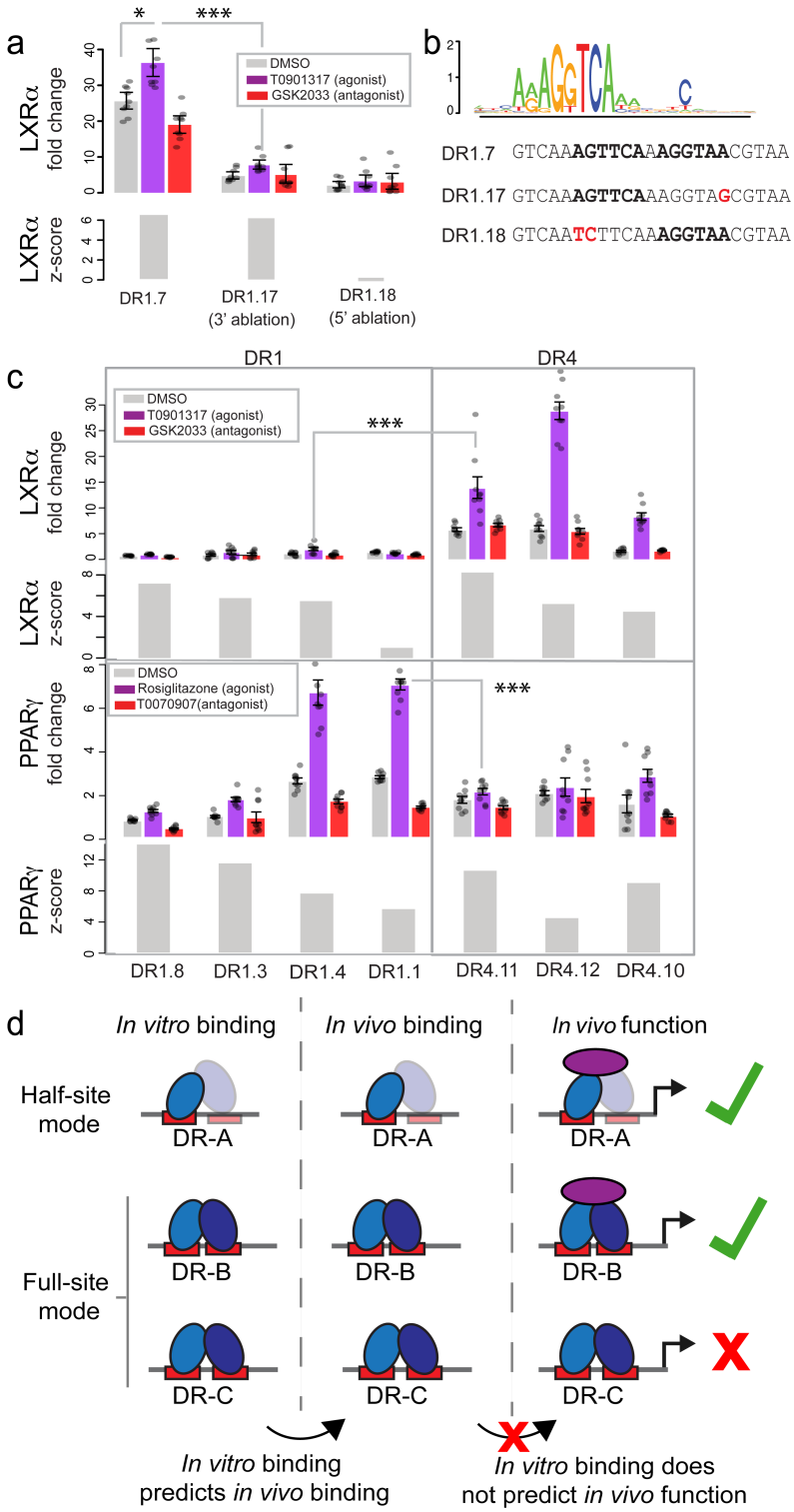
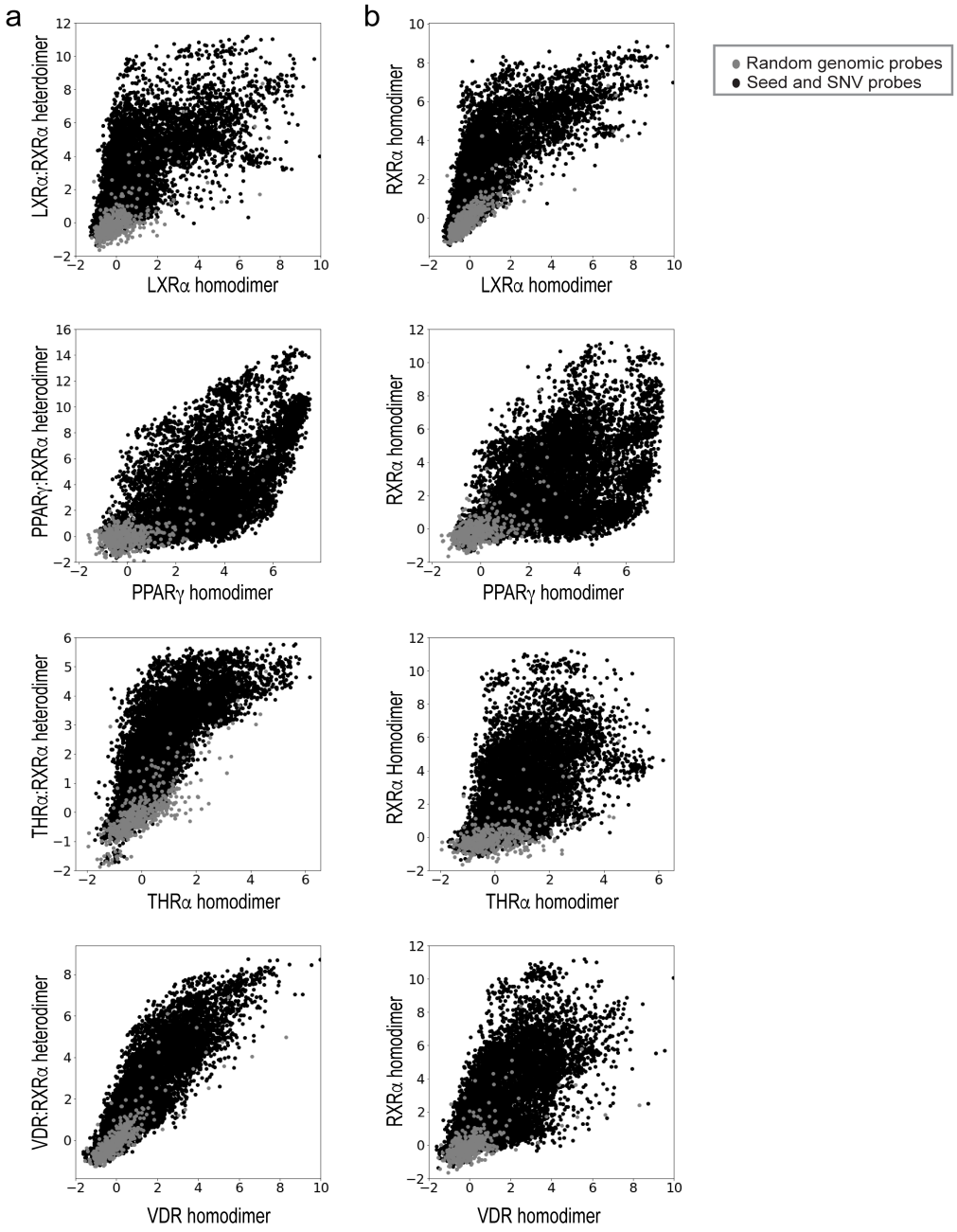


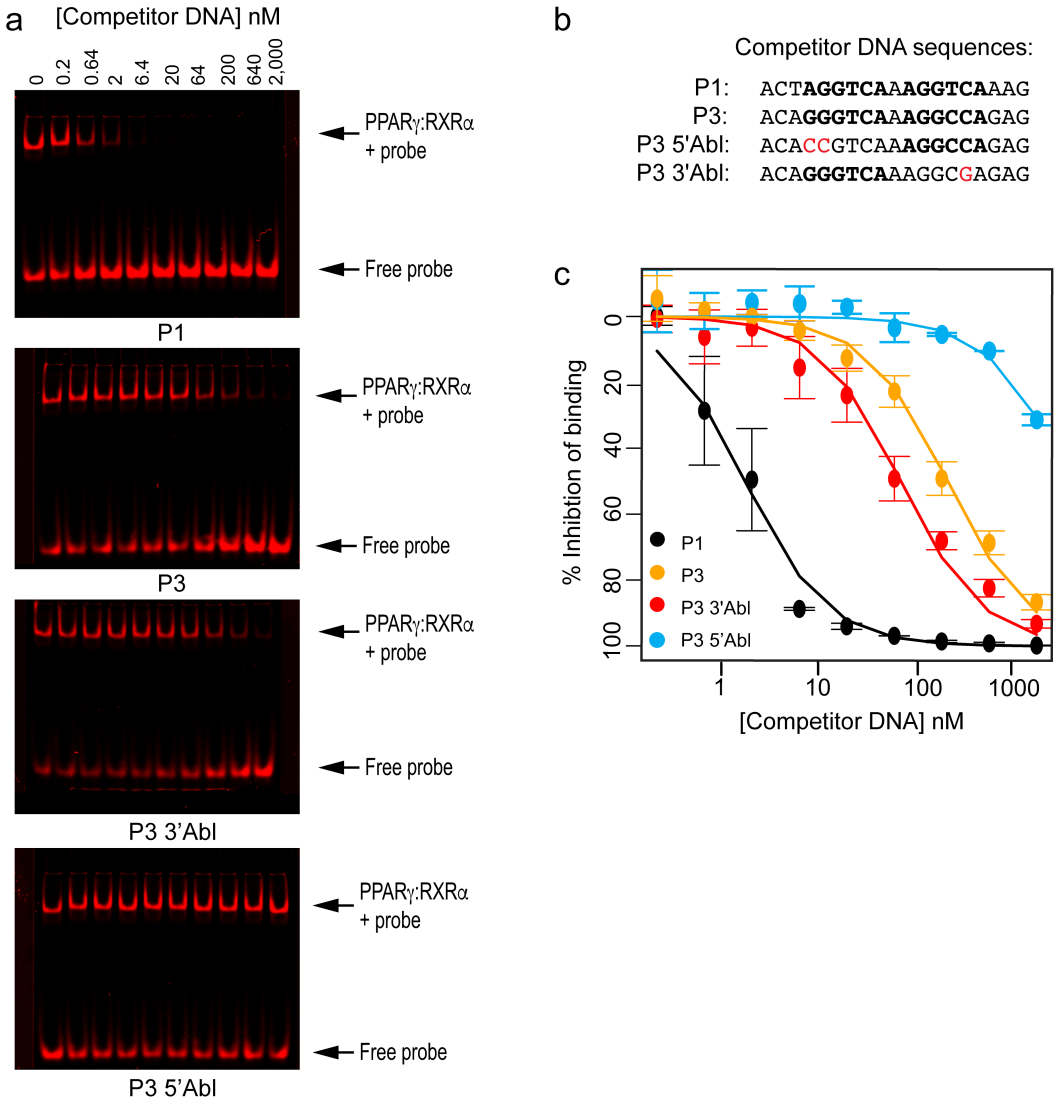
Figure 2.7: Activity versus affinity for distinct classes of NR-binding sites.

a LXRα-dependent activity and binding affinity of a sequence bound in a half-site mode. Luciferase reporter gene activation, and corresponding *z*-scores, are shown for the DR1.7 sequence, which is bound in a half-site mode on PBM, and sequences with each half-site ablated (DR1.17 and DR1.18), sequences shown in **2.7b**. Fold-change reporter expression indicates luciferase activity in HEK293T cells over-expressing LXRα and RXRα normalized to cells not over-expressing these proteins. Fold-change expression is shown for cells treated with DMSO (vehicle), agonist (T0901317), or antagonist (GSK2033), and values represent mean over nine replicate measurements (error bars = SEM). Reporter gene *p*-values: * < 0.01, *** < 0.0001 (calculated using Student's two-tailed *t*-test). **b** Logo for LXRα heterodimer binding to DR1.7, and sequences for DR1.7, DR1.17, and DR1.18 discussed in **2.7a**. **c** LXRα- and PPARγ-dependent activity and PBM-derived binding scores to select DR1 and DR4 sites. Fold-change expression for LXRα is as described in **a**. Fold-change for PPARγ is shown for cells treated with DMSO (vehicle), agonist (rosiglitazone), or antagonist (T0070907), and values represent the mean over nine replicate measurements (error bars = SEM). **d** Overview of relation between NR in vitro binding, in vivo binding, and function. Data and analysis by Ashley Penvose.



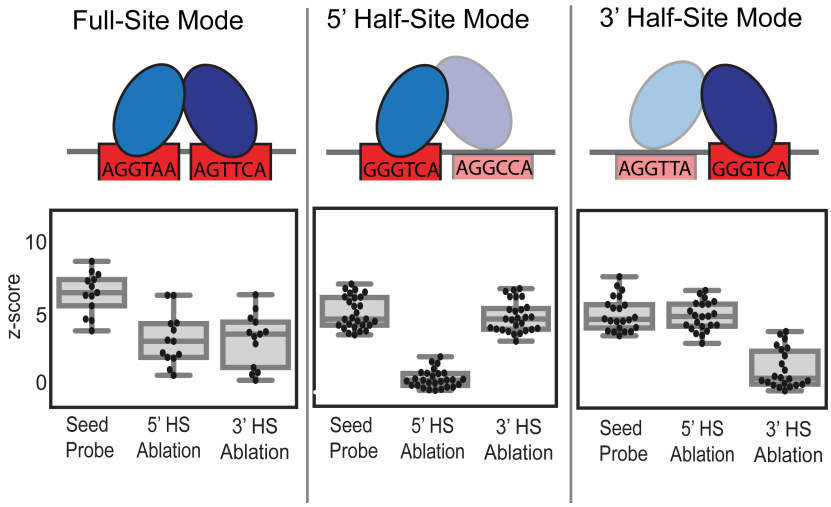
Supplementary figure 2.1: Comparison of NR homodimer and heterodimer binding.

Z-scores for **a** the indicated NR as a heterodimer with RXR against the corresponding NR homodimer or for **b** RXR homodimer against the indicated NR homodimer. Dots represent average over ~5 replicates for all 10,728 unique SNV probes and 500 background probes. Data and analysis by Ashley Penvose.



Supplementary figure 2.2: Competition EMSA experiments for PPARγ:RXRα

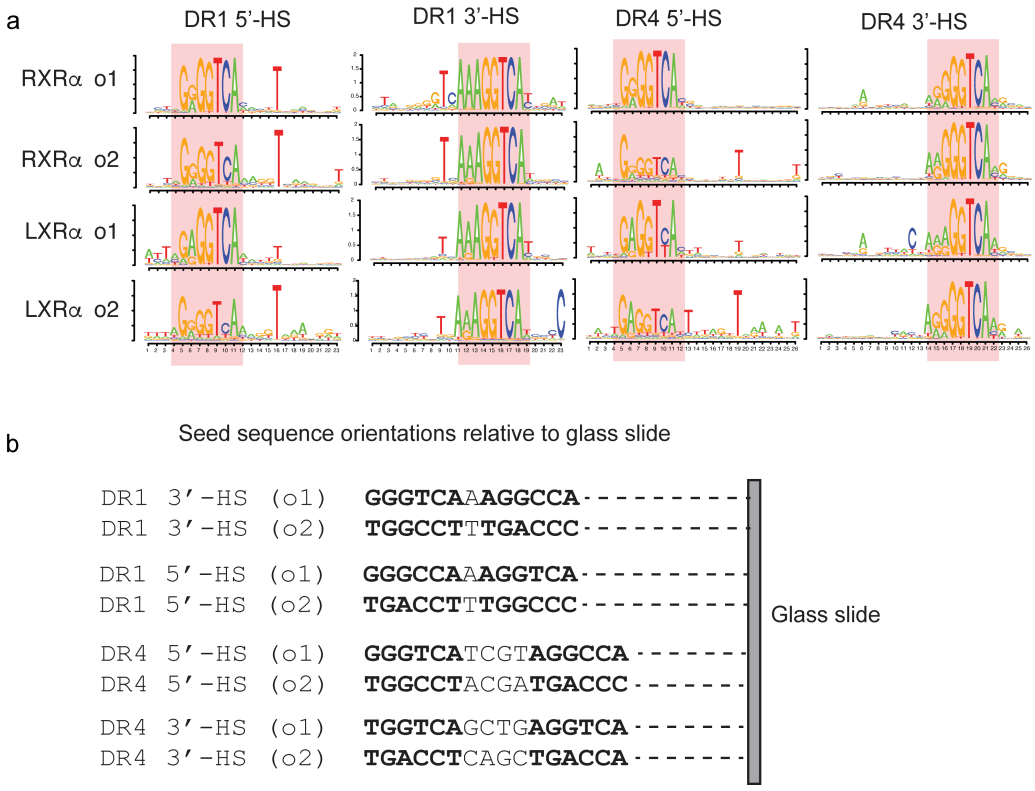
a Representative EMSA gels of competition for binding by PPARγ:RXRα to labeled DNA probe (P1, as described in Fig. 1g) and four unlabeled competitor DNA sequences whose sequence are shown in **b**. **c** Inhibition curves determined by quantifying the intensity of the bound probe band at different competitor concentrations for the different competitor experiments (error=STDEV, n = 2). Data and analysis by Ashley Penvose.



Supplementary figure 2.3: Impact of half-site ablation on LXRα binding.

This analysis is similar to the analysis shown in Fig. 2.3, but is expanded to include all seed sequences for which z-score >3. For all seed sequences for which LXRα bound with z-score > 3, the box and whiskers plots of the z-score distributions of binding in full-site mode and each half-site mode are shown, along with the corresponding 5' or 3' half-site ablations of each seed sequence. Center line: median; box limits: upper and lower quartiles; whiskers: last datum within 1.5x interquartile range. Data and analysis by Ashley Penvose.

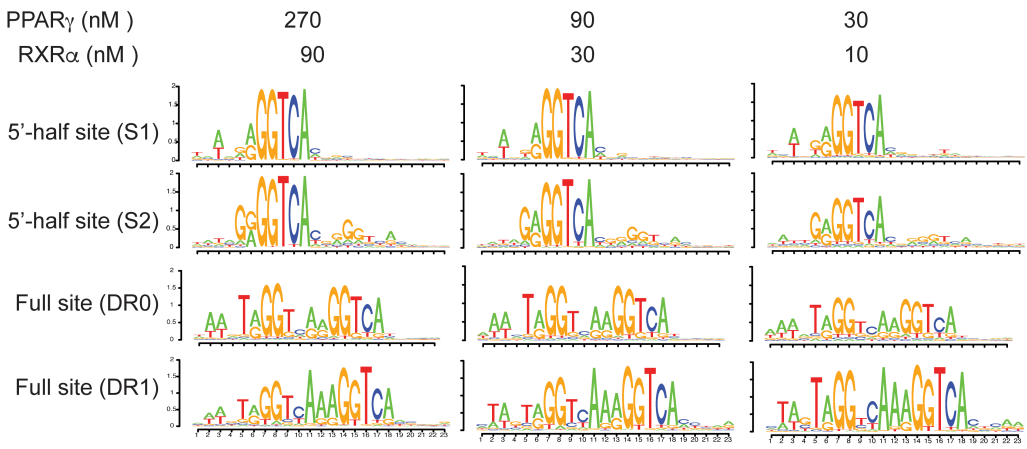
Supplementary figure 4



Supplementary figure 2.4: Impact of PBM probe orientation on NR binding logos.

a DNA binding logos for RXRα homodimers and LXRα:RXRα are shown for DNA sequences bound in either a 5' half-site or 3' half-site binding mode. DNA binding logos were determined separately from PBM probes in which the binding site (and all SNVs used in the logo determination) are oriented in either the o1 or o2 orientation with respect to the glass slide (schematized in **b**). Bases indicating the binding mode preference are highlighted with the red overlay box**. b** Schematic of DNA seed sequences used to generate the logos showing the orientation relative to the microarray glass slide. Data and analysis by Ashley Penvose in response to reviewer comments.

a

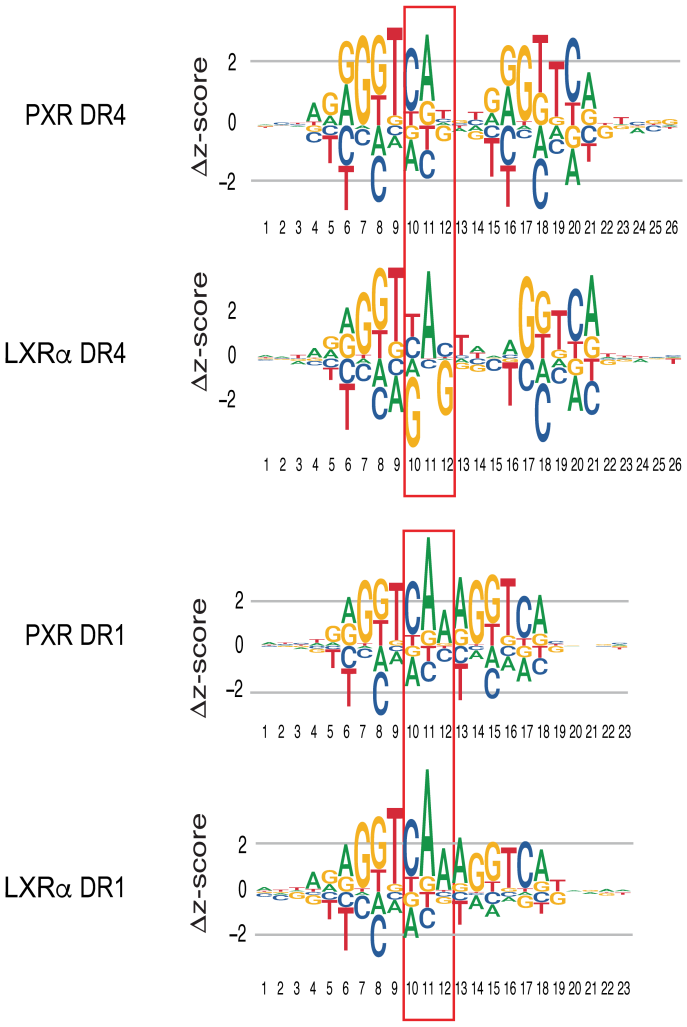


b Seed Sequences for Motif Models

5'-half site (S1)	GAACT**AGGTCA**CC**AGGACA**GTGAA
5'-half site (S2)	ATACA**GGGTCA**CT**AGGCCA**GAGTA
Full site (DR0)	TATGT**AGGTTAGGGTCA**TTCAG
Full site (DR1)	CTAGC**AGGCCA**A**GGGTCA**CTCAA

Supplementary figure 2.5: Impact of protein concentration on NR binding logos.

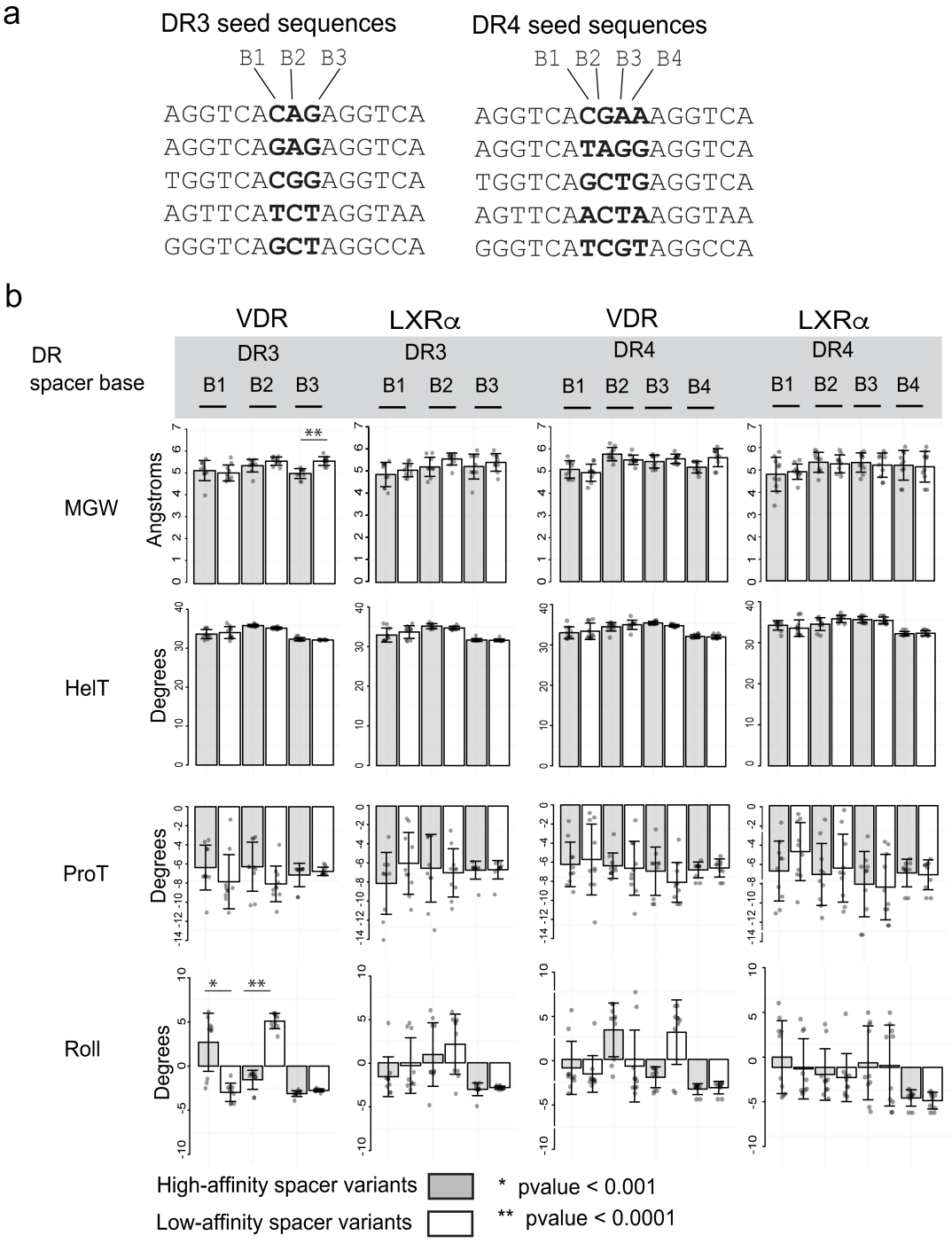
a PPARγ:RXRα DNA binding logos for DNA seed sequences bound in full or half-site binding modes are shown for PBM experiments performed at three different concentrations. The concentration of each monomer used in each PBM experiment is indicated. **b** The seed sequences for which the logos in **a** were generated. Identifiable DR half-sites in each binding sequence are shown in bold. Data and analysis by Ashley Penvose in response to reviewer comments.



Supplementary figure 2.6: DNA energy matrix logos for LXRα and PXR.

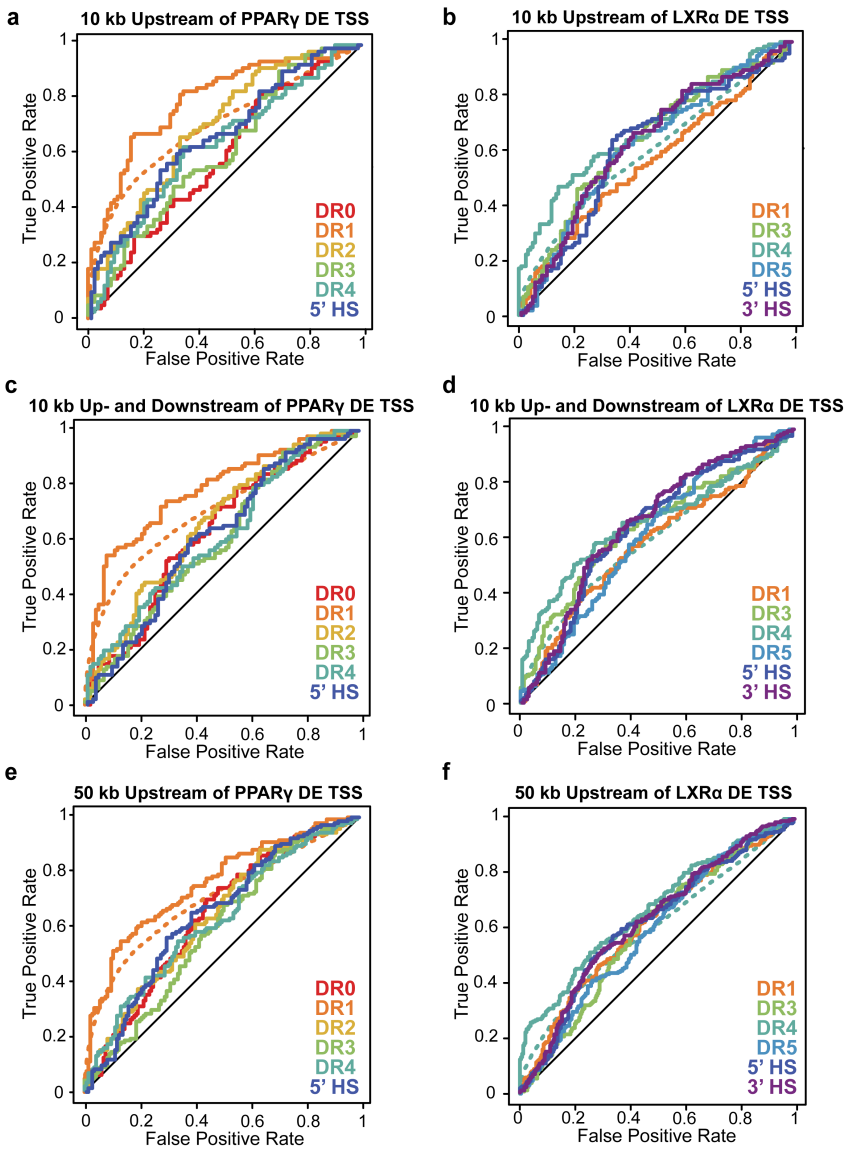
DR1 and DR4 logos, directly representing Δ z-scores of SNV binding, are shown for LXRα and PXR. DR4 logos are derived from the same experiments as those in Fig. 5 and are shown for comparison. Positive Δ z-scores indicate z-scores higher than the median z-score for all base variants at that position. Data and analysis by Ashley Penvose.

Supplementary figure 7



Supplementary figure 2.7: DNA-shape parameters of spacer sequences for high and low-affinity NR binding sites.

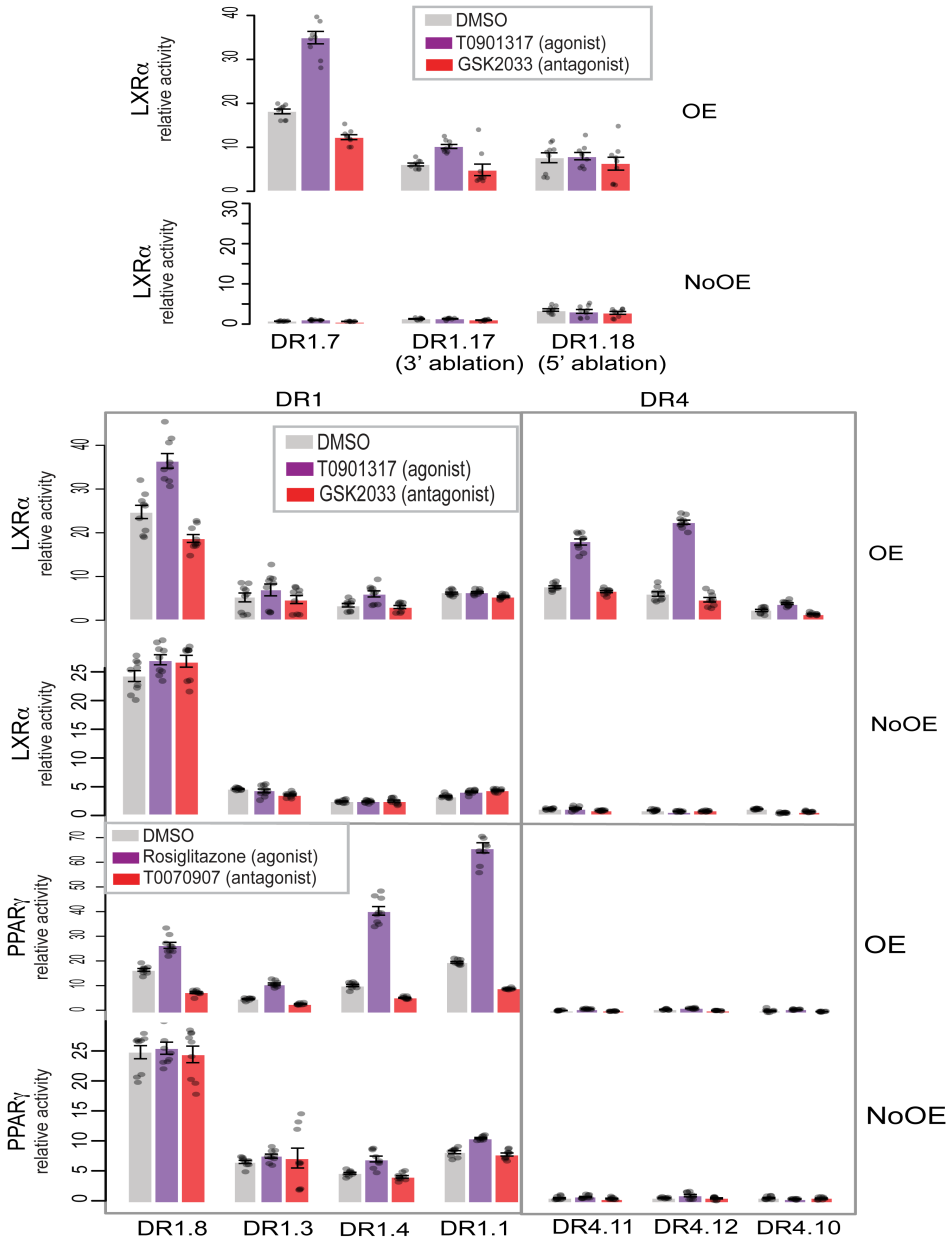
a Schematic of DNA seed sequences used to analyze DNA shape features (shown in **b**). Base positions in the spacer sequence between the DR half-sites are indicated in bold and referred to as B1,B2,B3 (DR3 site) and B1,B2,B3,B4 (DR4 site). Seed sequences were selected to represent diverse spacer sequences. **b** Distribution of DNA shape features for spacer sequences in either high-affinity sites (grey bars) or low-affinity sites (white bars). Data are shown for VDR and LXRα heterodimer binding experiments. For each of Supplementary figure 2.7: DNA-shape parameters of spacer sequences for high and low-affinity NR binding sites. For the 5 seed sequences (at each spacer length), we identified the two highest affinity and the two lowest affinity spacer sequence variants. Therefore, there are 10 (i.e, 5x2) high-affinity and 10 low-affinity spacer sequences considered for each bar plot. For each of the 10 spacer variants, DNA shape parameters were calculated at each base position using the TFBSshape server – major groove width (MWG), helix twist (HelT), propeller twist (ProT), and roll (Matys et al., 2006). Shown at each base position is the mean parameter over 10 sequences (error = STDEV). Distributions that were significantly different between the high and low-affinity sequences are shown (p-value calculated using a two-tailed t-test). Data was generated by Ashley Penvose. DNA shape analyses were performed by Trevor Siggers in response to reviewer comments.



Supplementary figure 2.8: Receiver operating characteristic (ROC) curves for PPARγ and LXRα motif enrichment in ChIP-seq data.

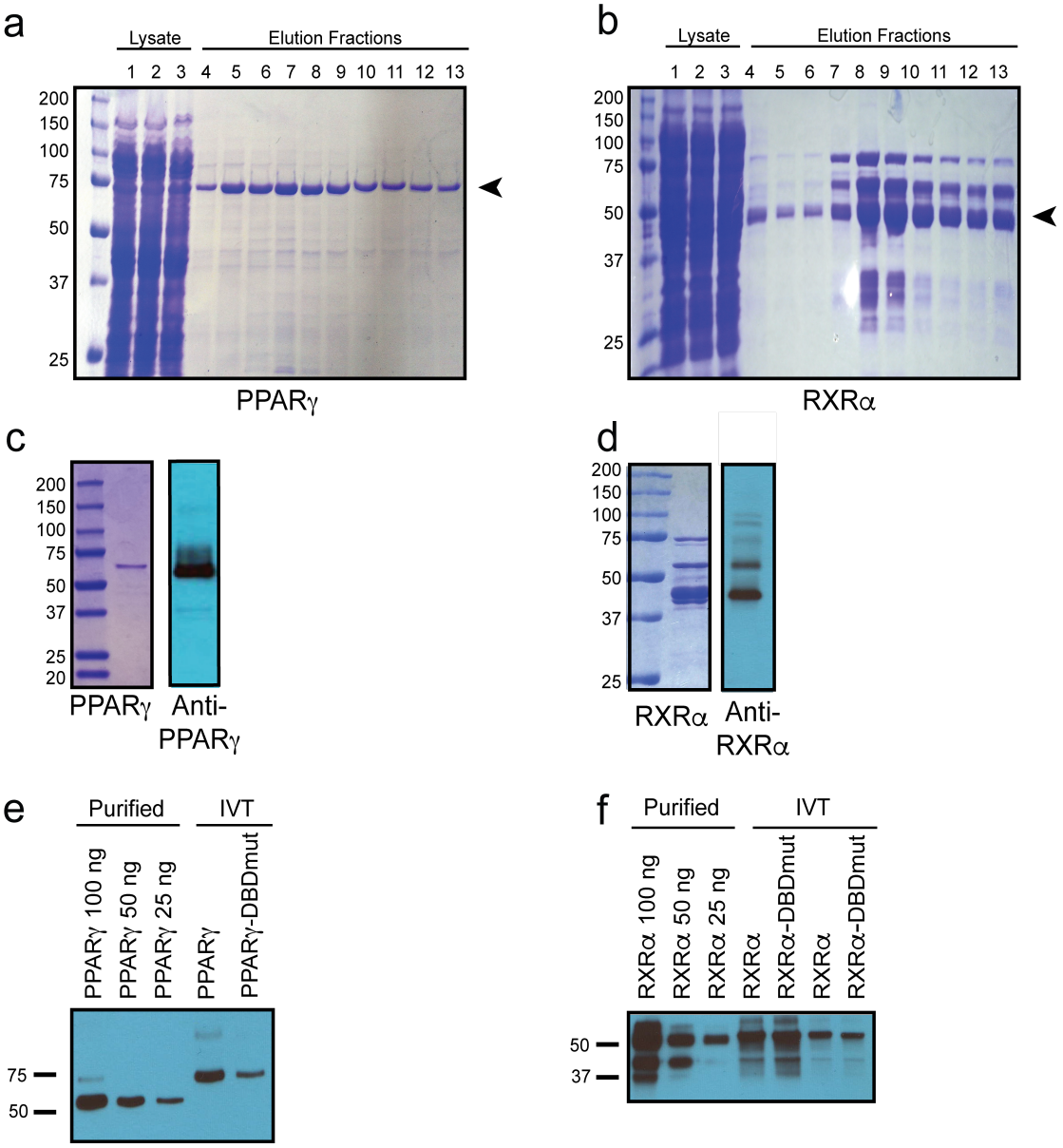
ROC curves for motif enrichment of PBM-derived PPARγ-binding models are shown for all reproducibly-bound PPARγ ChIP-seq peaks (dotted lines, **a,c,e**) and reproducibly-bound PPARγ ChIP-seq peaks occurring within **a** 10 kb upstream, **c** 10 kb upstream or downstream, and **e** 50 kb upstream of the transcription start site of differentially expressed genes (solid lines, Methods). ROC curves for motif enrichment of PBM-derived LXRα binding models are shown for all reproducibly-bound LXRα ChIP-seq peaks (dotted lines, **b,d,e**) and reproducibly-bound LXRα ChIP-seq peaks occurring within **b** 10 kb upstream, **d** 10 kb upstream or downstream, **f** and 50 kb upstream of the transcription start site of differentially expressed genes (solid lines, Methods). ROC curves determined using different PWMs for different DR and half-site (HS) modes are

indicated. The models tested in these analyses were generated by Ashley Penvose. The genomic enrichment analyses were performed by Jessica Keenan.



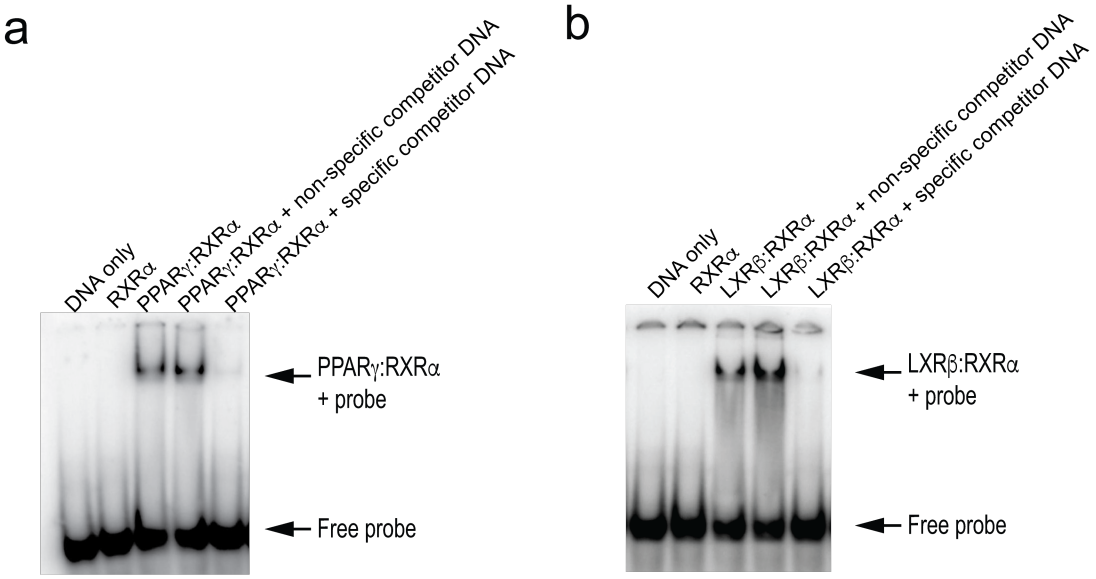
Supplementary figure 2.9: Impact of NR over-expression on reporter gene activity.

a,b LXRα- and PPARγ-dependent activity for the sequences described in Fig. 7 in the same treatment conditions. Shown separately are the luciferase activity values for the cells in which the NR:RXRα proteins were overexpressed (OE) and the values in which the proteins were not overexpressed (NoOE), each normalized to empty vector. Fold-change values in Fig. 7 are the ratio of these sets of values (i.e., OE/NoOE). Values represent the mean over nine replicate measurements (error bars = SEM). Data and analysis by Ashley Penvose.



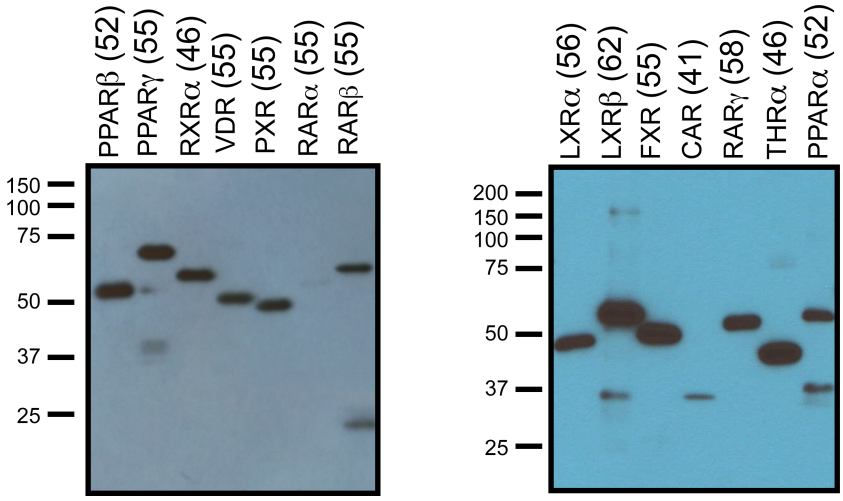
Supplementary figure 2.10: Protein quality control for

a,b SDS-PAGE of raw PPARγ and RXRα overexpression lysate from *E. coli* and the fractions eluted after His-Tag purification of PPARγ and RXRα. **c,d** SDS-PAGE and western blot of purified PPARγ and RXRα. **e,f** Anti-His western blot of purified and IVT-produced PPARγ, PPARγ-DBDmut, RXRα, and RXRα-DBDmut. Data and analysis by Ashley Penvose.



Supplementary figure 2.11: DNA binding activity test for purified proteins

Purified His-tagged PPARγ, LXRβ, and RXRα were incubated with 1nM of ^{32}P-labeled DNA probe in the presence or absence of 10x unlabeled competitor DNA of either the same sequence (specific competitor) or a different sequence to which the protein of interest should not show sequence-specific binding (non-specific competitor). Free probe and probe-protein complexes are indicated with arrows. Data and analysis by Ashley Penvose.



Supplementary figure 2.12: Quality control for purchased proteins

Example Anti-His western blots of the purchased proteins that are listed in Table 2.1. Predicted molecular weight of each protein (kDa) is indicated in parentheses next to the protein name. Note 1: The RARα and CAR samples shown here were not used on the PBM as they did not appear to be present at concentration reported by the manufacturer. A replacement sample of RARα was obtained and appeared to be the correct concentration. The manufacturer could not provide a replacement CAR protein. Note 2: The PPARβ protein shown here was tested on PBM and did not show DNA binding activity. Data and analysis by Ashley Penvose.

Protein	Source	Product Number	Tag	Concentration used on PBM (nM)	R^2 of antibody replicates on PBM
CAR	Protein One	P1096-01	His	N/A	N/A
FXR	Protein One	P1044-01	His	270	0.95
LXRα	Protein One	P1045-01	His	270	0.94
LXRβ	Protein One	P1046-02	His	270	N/A
LXRβ	Abcam	ab81924	His	270	0.93
PPARα	Protein One	P1048-01	His	270	N/A
PPARα	Abcam	ab81925	GST	270	0.97
PPARβ	Protein One	P1049-01	His	270	N/A
PPARγ Isoform 1	Purified in our lab	N/A	His	180***	0.98
PPARγ Isoform 1	Protein One	P1050-01	His	270	0.97
PXR	Protein One	R1082-02	His	270	0.91
RARα	Protein One	P1054-01	His	270	0.94
RARβ	Protein One	P1056-01	His	270	0.98
RARγ	Protein One	P1055-02	His	270	0.92
RXRα	Purified in our lab	N/A	Cleaved His	90(Heterodimer)*** 180(Homodimer)***	0.98
RXRα	Active Motif	81082	His	90***	0.97
THRα	Protein One	P1052-02	His	90***	0.97
VDR	Protein One	R1084-2	His	270	0.91
RXRα wild-type	IVT in our lab	N/A	His	90*	0.98

RXRα DBDmut	IVT in our lab	N/A	His	90*	0.89
PPARγDBDmut	IVT in our lab	N/A	His	270	N/A
PPARγ wild-type	IVT in our lab	N/A	His	270	0.84

Table 2.1 Sources and concentrations for proteins used in all experiments

All proteins used were generated from human coding sequences. *** Indicates that a concentration other than 270 nM was used in this experiment. A value of N/A for the protein concentration or R^2 of antibody replicates means that the proteins did not pass quality control checks by western blot or DNA binding activity experiments.

Response Element Sequence (5'→3')	Site Type	NR	Source
CAAAACTAGGTCAAAGGTCA	DR1	PPAR	Juge-Aubry 1997
ACAGTTCATGAAGTTCATC	DR1	CAR, PXR	Xie 2000
CAAAACTAGGTCAAAGGTCACCTGC	DR1	PPAR (all)	Juge-Aubry 1997
AAAAACTGGGCCAAAGGTCTCAGAA	DR1	PPAR (all)	Juge-Aubry 1997
CAAATATAGGCCATAGGTCAGTGAT	DR1	PPAR (all)	Juge-Aubry 1997
TGAAACTAGGGTAAAGTTCAGTGAG	DR1	PPAR (all)	Juge-Aubry 1997
CAAATGTAGGTAATAGTTCAATAGG	DR1	PPAR (all)	Juge-Aubry 1997
GCATTCTGGGTCAAAGTTGATCCCC	DR1	PPAR (all)	Juge-Aubry 1997
ACTCCCACGGCCAAAGGTCATGAGA	DR1	PPAR (all)	Juge-Aubry 1997
CTGAACTAGGGCAAAGTTCACCTGC	DR1	PPAR (all)	Juge-Aubry 1997
CACAACTGGGATAAAGGTCTCGCTG	DR1	PPAR (all)	Juge-Aubry 1997
GGGGACCAGGACAAAGGTCACGTTC	DR11	PPAR (all)	Juge-Aubry 1997
GCGTTACAGGACAAAGGCCACCGAG	DR1	PPAR (all)	Juge-Aubry 1997
TCTTACTGGATCAGAGTTCACTAGT	DR1	PPAR (all)	Juge-Aubry 1997
AGGGCGCTGGGCAAAGGTCACCTGC	DR1	PPAR (all)	Juge-Aubry 1997
GAGAGCAAGGTAGAAGGTCAAGAAA	DR1	PPAR (all)	Juge-Aubry 1997
GTCTTTCAGGGCAACAGTCACATGC	DR1	PPAR (all)	Juge-Aubry 1997
TCTCTCTGGGTGAAATGTGCATTTC	DR1	PPAR (all)	Juge-Aubry 1997
GTGTTAGAGGGCACAGGTCCAGTGG	DR1	PPAR (all)	Juge-Aubry 1997
CAACTAGGTCATCAGGTCA	DR2	PPARα	Juge-Aubry 1997
GCTTCAGAACACCAGGAGAACAGAGAG	DR5mut	RAR	Umesono 1991
AGGTGAACAGGAGGACA	DR5mut	RAR	Umesono 1991
AGGTGAGGCTGCGGTGA	DR5mut	RAR	Umesono 1991
TTAAGGGTTCACCGAAAGTTCACTCGC	DR5	RAR	Umesono 1991
GCTTCAGGTCACCAGGAGGTCAGAGAG	DR5	RAR	Umesono 1991
AGGTCACCAGGAGGTCA	DR5	RAR	Umesono 1991
AGGTGAACAGGAGGTCA	DR5	RAR	Umesono 1991

GGTTCACCGAAAGTTCA	DR5	RAR	Umesono 1991
AGGTCACTGACAGGGCA	DR5	RAR	Umesono 1991
GGGTCATTCAGAGTTCA	DR5	RAR	Umesono 1991
TAAGGTCAATCCGAAGTCACTC	DR5	RAR	Umesono 1991
TAAGGTCAAGCAGTGGTCCCTC	DR5	RAR	Umesono 1991
TAAGGTCAATAAGGGGTGACTC	DR5	RAR	Kurokawa 1993
TAAGGTCAAGGAAAGGACAATC	DR5	RAR	Kurokawa 1993
TAAGGTCAACGATAGGGCACTC	DR5	RAR	Kurokawa 1993
TAAGGTCAACATAGGGTCTCTC	DR5	RAR	Kurokawa 1993
TAAGGTCAAAGGGCGGACACTC	DR5	RAR	Kurokawa 1993
GGTAGGGTTCACCGAAAGTTCACTCGA	DR5	RAR	Kurokawa 1993
CTTAAGGTCAACGGGTGGGCACTCAG	DR2	RAR	Kurokawa 1993
CTTAAGGTCAAGAGTTCAATCCTCAG	DR2	RAR	Kurokawa 1993
CTTAAGGTCAAAAGGTCGTTGCTCAG	DR2	RAR	Kurokawa 1993
CTTAAGGTCAAAAGTTGACGCCTCAG	DR2	RAR	Kurokawa 1993
TAGCTAGGTAAGATCAGGTAAGTAGC	DR4	RAR & THR	Umesono 1991
GGGTAAGGTCACAATAAGGTCACGAAG	DR5	THR	Kurokawa 1993
GCTGGAGGTGACAGGAAAACAGCAAG	DR4mut	THR	Umesono 1991
GCTGGAAATGACAGGAGGACAGCAAG	DR4mut	THR	Umesono 1991
GCTTCAGAACACAGGAGAACAGAGAG	DR4mut	THR	Umesono 1991
TGGTCACAGGTGGTCA	DR4mut	THR	Umesono 1991
AGGTAAGATCAGGTAA	DR4mut	THR	Umesono 1991
GGGTGAATGGGGGTGA	DR4mut	THR	Umesono 1991
GGTTCACCGAAGTTCA	DR4mut	THR	Umesono 1991
AGCTGGGGTTAGGGGAGGACAGTAAG	DR4	THR	Umesono 1991
GCTGGAGGTGACAGGAGGACAGCAAG	DR4	THR	Umesono 1991
GCTGGAGGTGACGAAAGGACAGCAAG	DR4	THR	Umesono 1991
GCTCAGGGTCATTTCAGGTCCTTGAA	DR4	THR	Umesono 1991
GCTTCAGGTCACAGGAGGTCAGAGAG	DR4	THR	Umesono 1991
AGGTCACAGGAGGTCA	DR4	THR	Umesono 1991
AGGTGACAGGAGGTCA	DR4	THR	Umesono 1991

AGGTGACAGGAGGACA	DR4	THR	Umesono 1991
GGGTTAGGGGAGGACA	DR4	THR	Umesono 1991
GGGTCATTTCAGGTCC	DR4	THR	Umesono 1991
GGGCCAGCTGAGGTTA	DR4	THR	Umesono 1991
AAGGGGATCCAGCTTCAGGTCACAGGAGGT CAGAGAGCT	DR4	THR	Meier 1993
GGGTAAGGTCAAATAAGGTCACGAAG	DR4	THR	Kurokawa 1993
TCTGGAGGTGACAGGAGCACAGCGGA	DR4	THR	Kurokawa 1993
TGATCAGGTCATCAGGTCACAGAT	DR2	THR	Kurokawa 1993
GCTTCAGAACAAGGAGAACAGAGAG	DR3mut	VDR	Umesono 1991
AGGTCAGCTAGGTTA	DR3mut	VDR	Umesono 1991
GCTTCAGGTCAAGGAGGTCAGAGAG	DR3	VDR	Umesono 1991
AGGTCAAGGAGGTCA	DR3	VDR	Umesono 1991
GGGTGAATGAGGACA	DR3	VDR	Umesono 1991
GGGTGAACGGGGGCA	DR3	VDR	Umesono 1991
GGTTCACGAGGTTCA	DR3	VDR	Umesono 1991

Table 2.2 Literature curated response elements

Response Element Sequence (5'→3')	Label in figure
CAAAACTAGGTCAAAGGTCA	Fig1e-Consensus
TGAAACTAGGGTAAAGTTCAGTGAG	Fig 1e-High Affinity DR1
ACTCCCACGGCCAAAGGTCATGAGA	Fig1e-Medium Affinity DR1
GAGAGCAAGGTAGAAGGTCAAGAAA	Fig1e-Low Affinity DR1

Table 2.3 DR1 sequences with EMSA-derived affinity ranks used as controls on PBM

Note: All sequence were obtained from (Juge-Aubry et al., 1997)

DNA sequence (5'→3')	DR site type	Label in figure
GCCAAACTAGGTCAAAGGTCAAAGTCGAGCCAACCA	DR1	Fig 1g- P1
GCTCTAGTTGGTCAAAGGTCACACACTCGAAATGCG	DR1	Fig 1g- P2
GCTCTTGTAGATCAAAGGTTAAGAAAGCCGAAGCCT	DR1	Fig 1g- P4
GCTATACAGGGTCAAAGGCCAGAGTACGAATGTCGA	DR1	Fig 1g- P3
GCTATACACCGTCAAAGGCCAGAGTACGAATGTCGA	DR1mut	Fig 1g- P3 5'Abl
GCTATACAGGGTCAAAGGCGAGAGTACGAATGTCGA	DR1mut	Fig 1g- P3 3'Abl
GCCCAGACAGGTGAAAGGTAACTCAGTCTGGTACGC	DR1	Fig 1g- P5 Abl
GCTCTTGTAGATCAGCAGGCTGAGAAAGCTGGCCAA	none	Fig 1g- P6

Table 2.4 DNA sequences from PBM used for competition EMSA experiments

DNA Sequence (5'→3')	Site Type	Label in figure
GCTCTAGTTGGTCAAAAGTCACACACTCGAAATGCG	DR1	Fig 7-DR1.1
GCTCTAGTTGGTCAAAGGTCACACACTCGAAATGCG	DR1	Fig 7-DR1.3
GCCGTCCAAGGTCACAGGTCAGTGAAGAGATGACGG	DR1	Fig 7-DR1.4
GCTGTCAAAGTTCAAAGGTAACGTAAGTATCCACAC	DR1	Fig 7-DR1.7
GCCAAACTAGGTCAAAGGTCAAAGTCGAGCCAACCA	DR1	Fig 7-DR1.8
GCTGTCAAAGTTCAAAGGTAGCGAAAGTATCCACAC	HS	Fig 7-DR1.17, 3' ablation of DR1.7
GCTGTCAATCTTCAAAGGTAACGTAAGTATCCACAC	HS	Fig 7-DR1.18, 5' ablation of DR1.7
GCCAAACTAGGTCACGAAAGGTCTAAGTCGAATGAC	DR4	Fig 7-DR4.10
GCCAAACTAGGTCACGAAAGGTCAAAGTCGAATGAC	DR4	Fig 7-DR4.11
GCCCTACTGGGCCACGAAAGGTCATTGATTCAGTGA	DR4	Fig 7-DR4.12
GCTCTTGTAGATCAGCAGGCTGAGAAAGCTGGCCAA	N/A	Insert for "empty vector" to mimic sequence context of test element inserts. Used for normalization to "empty vector"

Table 2.5 DNA sequences used in reporter elements

	DR0	DR1	DR2	DR3	DR4	DR5
Seed 1	AAACTAGGTCAAGGTCAAAGT	A	AG	CAG	CGAA	CCGGA
Seed 2	CTAGTTGGTCAAGGTCACACA	A	GG	CGG	GCTG	GTAAA
Seed 3	CTCGGAGGACAAGGTCAGAGC	A	GA	CGA	CGAA	ACAGG
Seed 4	CTACTGGGCCAAGGTCATTGA	A	TC	AAA	CGAA	TTCAG
Seed 5	CTTACGGGTCAAGGTTATAGC	A	CG	ACG	CTCC	CTGAC
Seed 6	CTAGCAGGCCAGGGTCACTCA	A	GA	TTT	ATCT	AGTTT
Seed 7	CATATAGGTAAAGGTGAAAGT	A	AC	TAC	TACA	ATGCT
Seed 8	CTTGTAGATCAAGGTTAAGAA	A	GC	AGC	ATAG	ACAGG
Seed 9	TCTCTAGGTAAAGTTCAGTGA	A	CA	CCA	AGAT	GACTC
Seed 10	TCGAAAGGTCGAGGTGAGTAT	A	TG	ACG	GTTG	TGCTC
Seed 11	CTATTAGTTGAAGGGCAACCG	A	TT	CTT	GACG	CCATA
Seed 12	CGACTGGGGCAAGGTTACCAT	A	TA	GAA	AAGC	TCAGC
Seed 13	GTCCAAGGTCAAGGTCAGTGA	A	GC	GAG	TAGG	ATGCT
Seed 14	CTGCTAGGTCATGGTCAGCTA	A	GT	GGT	ACGA	ACTGA
Seed 15	GAACTAGGTCAAGGACAGTGA	A	CC	TCC	GAAC	TTGAC
Seed 16	AGCCTAGGTCAGGGCCAGTAT	A	TT	TGC	GATC	TTCAG
Seed 17	TATGTAGGTTAGGGTCATTCA	A	TC	AGG	GCTG	CATCC
Seed 18	ATACAGGGTCAAGGCCAGAGT	A	CT	GCT	TCGT	GCTAG
Seed 19	CAGACAGGTGAAGGTAACTCA	A	GA	TGA	GTGA	AACGA
Seed 20	GAATGAGGTTAAGATCAGGGA	A	TA	TTA	CCAA	TACGT
Seed 21	GTCAAAGTTCAAGGTAACGTA	A	CT	TCT	ACTA	GGACG
Seed 22	TACCAAGGTGAAGGTCGTTGA	A	AT	ACA	GCTG	eTCGTG
Seed 23	GAACGAGGGCAAGTTGATAGC	A	TG	ATG	TTCG	CGCTG
Seed 24	CGAGTAGGTTAGGGGCATTCA	A	CG	GTC	CGAG	CCAAT

Table 2.6 DNA seed sequences for custom-designed NR PBM

Shown are the 24 starting seed sequences with no intervening spacer bases (DR0) and the bases of the spacer sequence that was inserted between the two half-sites to create the corresponding seed at longer spacer lengths (DR1-5). For each seed, the first half-site of the DR is highlighted in red and the second half-site of the DR is in blue. The 5 nt of each the 5' and 3' flanks are shown in black. On the PBM, SNV probes of the seed sequences shown above were included across the entire region shown.

NR	Assay
FXR	PBM
LXRα	PBM, Luciferase reporter, Genomic enrichment analyses
LXRβ	PBM
PPARα	PBM
PPARγ	PBM, competition EMSA, PBMs with DBDmut proteins, Luciferase reporter, Genomic enrichment analyses
PXR	PBM
RARα	PBM
RARβ	PBM
RARγ	PBM
RXRα	PBM, PBMs with DBDmut protiens
THRα	PBM
VDR	PBM

Table 2.7 Summary of assays done on each nuclear receptor

CHAPTER THREE: TYPE II NR DNA BINDING SPECIFICTY AND COREGULATOR RECRUITMENT FROM NUCLEAR LYSATE

All experiments were performed by Ashley Penvose. 3T3-L1 cell culture and differentiation for some samples was performed by Stephanie Kim from the lab of Jennifer Schlezinger at Boston University Medical School. Data analysis, figures, and writing by Ashley Penvose as advised by Trevor Siggers.

3.1 Abstract

In Chapter 2, I characterized the binding of purified type II NRs and demonstrated that most NRs bind with high affinity to many different DR spacer lengths. This finding disagrees with the prevailing model in which NRs exhibit binding preferences for a particular DR spacer length. However, using reporter gene assays, I also demonstrated that NR-dependent gene expression is mediated by binding sites that conform more closely to canonical models of NR DR-spacer preference. These observations suggest that (1) NR-DNA binding is promiscuous, but NRs can only regulate gene expression when bound to select DR sites, and/or (2) NR-DNA binding in cells is more specific than was observed with purified proteins. To test this second hypothesis, I utilized nextPBM to measure NR binding from 3T3-L1 adipocyte nuclear lysate. I observe high-affinity binding of RXRα-containing complexes to a broad range of DR spacer lengths (DR0-5). Interestingly, I find that the binding preferences of PPARγ from nuclear lysate are altered compared to purified PPARγ:RXRα, and better recapitulate the previously reported

preference for DR1 binding. To explore the possibility that DNA sequence may act as an allosteric regulator of coregulator assembly and NR function, I measured the recruitment of SRC1 from 3T3-L1 adipocyte nuclear lysates with nextPBM. I find that SRC1 recruitment shows strong DNA specificity to sites that resembles a NR half-site, but also show an extended sequence preference for the 3 bases that are 5' of the NR-like half-site.

3.2 Introduction

Nuclear receptors (NRs) are a large family of ligand-activated transcription factors (TFs) that serve as key regulators of development, metabolic homeostasis, inflammation, and protection from xenobiotics (Evans and Mangelsdorf, 2014). NRs rely on sequence-specific DNA binding to regulate their target genes (Evans and Mangelsdorf, 2014). NRs bind the sequence 5'-RGKTCA-3' organized as direct repeats (DRs) with a variable length spacer of 0–5 base pairs (bp) (DR0-DR5) (Weikum et al., 2018). Previous studies had suggested that each NR has distinct DNA binding preferences for certain DR spacer lengths, which enable individual NRs to distinguish their target genes (Juge-Aubry et al., 1997; Umesono et al., 1991). However, our recent work demonstrated that purified NRs bind with high affinity to many different DR spacer lengths, and have largely overlapping binding specificity (Penvose et al., 2019). Using reporter assays, I also found that canonical DR spacer preferences better predict NR function than NR binding (Penvose et al., 2019). These observations suggest that (1) NR-DNA binding is promiscuous, but NRs can only regulate gene expression when bound to select DR sites, and/or (2) NR-DNA binding in cells is more specific than was observed with purified proteins.

Recruitment of coregulatory proteins (CoRs) to TFs is an essential step in transcriptional regulation (Rosenfeld and Glass, 2001). Preferential assembly of distinct transcriptional complexes at different regulatory loci is thought to underlie locus-specific transcriptional responses (Zhang and Verdine, 1999). Previous work has shown that DNA can act as an allosteric regulator of co-regulator assembly for the NRs (Hall et al., 2002) and that allostery can impact NR function (Meijsing et al., 2009).

As key sensors of metabolic state and regulators of metabolic gene programs, NRs are central to the function of adipocytes (Jacobi et al., 2012). Highlighting the importance of NRs in adipocyte function, PPARγ is a master regulator of adipogenesis and is both necessary and sufficient for the conversion of pre-adipocytes into mature adipocytes (Rosen et al., 1999). Many NRs (30 of 49) are expressed in 3T3-L1 adipocytes including PPARβ and γ, LXRα and β, TRα and β, RARα, β, and γ, and RXRα, β and γ (Fu et al., 2005). As such, 3T3-L1 adipocytes represent an ideal cell-type to study NR-DNA binding specificity in a cellular context.

To test whether NR-DNA binding preferences are altered in a cellular context, we performed nextPBM experiments using nuclear lysate from 3T3-L1 adipocytes to measure the binding of NRs. We are able to capture high-affinity binding of RXRα-containing complexes to a broad range of DR spacer lengths (DR0-5). Interestingly, we find that the binding preferences of PPARγ from nuclear lysate are altered compared to purified PPARγ:RXRα, and better recapitulate the previously reported preference for DR1 binding (Juge-Aubry et al., 1997). To explore the possibility that DNA can act as an

allosteric regulator of coactivator recruitment, we measured SRC1 recruitment from 3T3-L1 nuclear lysate on nextPBM. I find that SRC1 recruitment shows strong DNA specificity that resembles a NR-like half-site with extended sequence preferences for the 3 bases 5' of the NR-like half-site.

3.3 Results

3.3.1 Measuring NR binding from 3T3-L1 lysate

Our recent work, in which we characterized the binding of purified type II NRs (Penvose et al., 2019), demonstrated that most NRs bind with high affinity at many different DR spacer lengths (Penvose et al., 2019). This finding disagrees with the prevailing model in which NRs exhibit binding preferences for particular DR spacer length (e.g., that PPARγ binds preferentially to DR1 sites) (Evans and Mangelsdorf, 2014; Palmer et al., 1995). To test whether NR binding in a cellular context is different than for purified proteins, I analyzed NR binding using our recently described nuclear extract protein binding microarray (nextPBM) approach (Mohaghegh et al., 2019), which measures TF binding using nuclear lysates instead of purified samples to better simulate the cellular environment. NextPBM has the advantage of querying TF binding in a cell-type specific context that accounts for endogenous post-translational modifications, and the presence of other transcription factors and coregulators (CoRs) that may alter the binding landscape of transcription factors. I used nextPBM to measure NR binding from 3T3-L1 adipocyte nuclear lysates (schematized in Fig. 3.1a**)**.

To examine type II NR binding by nextPBM, I utilized the same PBM microarray design that we used previously to characterize the binding of purified type II NRs (Penvose et al., 2019). Briefly, this PBM design contains over 1600 unique sequences at each of six DR spacer lengths (DR0-DR5). For each DR spacer length, I measured NR binding to 24 starting sequences, which we refer to as seed sequences (Fig. 2.1c). Seed sequences were generated by combining different half-site sequences exhibiting a range of degeneracy from the consensus 5'-RGKTCA-3' (Penvose et al., 2019). I also measure binding to all single-nucleotide variants (SNVs) of each seed sequence. This SNV-based design allows me to generate binding models for each seed sequence (described in detail in (Penvose et al., 2019)).

To gain a broad picture of the NR binding landscape from 3T3-L1 adipocyte lysate, I performed nextPBM experiments using an antibody against RXRα, which allows me to detect binding of complexes containing RXRα, including RXRα homodimers and all type II NRs as heterodimers with RXRα (schematized in Fig. 3.1b). To quantify binding specificity, nextPBM fluorescence values were converted into z-scores using a set of 500 random genomic sequences (Methods 3.5.4). We observe high specificity NR binding over a range of affinities, up to a z-score of 12 (Fig. 3.1c). Comparing independent biological samples of 3T3-L1 adipocytes, I observe strong agreement in binding of RXRα (Fig. 3.1d, R^2=0.93). These results demonstrate that using nextPBM I can reproducibly capture high specificity binding of RXRα complexes from nuclear lysates generated from 3T3-L1 adipocytes.

To examine the broad DNA-binding preferences of RXRα complexes from 3T3-L1 adipocyte nuclear lysate, we examined which DR spacer lengths are bound with high-affinity by RXRα complexes. We detected high-affinity binding of RXRα to all DR spacer lengths examined, DR0-5 (Fig. 3.1e). In our previous study of purified NR heterodimers (Penvose et al., 2019) we used a SNV modeling approach (Andrilenas et al., 2018; Penvose et al., 2019) to generate binding logos. The SNV approach allowed us to define the binding mode of the NR dimers to each individual DR seed site. Using this same approach I annotated the binding mode of RXRα dimers to all high-affinity sites as either full-site (i.e., both DR half-sites are engaged) or half-site (i.e., only a single DR half-site is engaged). I observe that almost all high-affinity RXRα binding events in our assay occur via a full-site binding mode, and that there appears to be an enrichment of high-affinity binding to DR1 and DR4 sites (Fig. 3.1e, blue dots). Examining the binding logos generated for our highest affinity DR sites at each spacer length (e.g., DR0-5), I observe base preferences consistent with the canonical 5'-RGKTCA-3' half-site preferences we previously defined using purified NR proteins (Fig. 3.1f). These results demonstrate that in 3T3-L1 adipocyte extracts there are RXRα dimers capable of binding with high-affinity to all DR spacer lengths (DR0-5) in a full-site mode.

3.3.2 NR binding landscape is altered in a cellular context

PPARγ is a key regulator of differentiation and maintenance of adipocytes, with a reported preference for DR1 DNA binding sites (Palmer et al., 1995, Juge-Aubry et al., 1997). To examine whether PPARγ from 3T3-L1 adipocyte lysates (PPARγ-AL) exhibit the same promiscuous DNA-binding preferences that I observed for purified

PPARγ:RXRα, I performed nextPBM experiments using 3T3-L1 adipocyte nuclear lysates and an antibody against PPARγ. A DNA binding logo generated for PPARγ-AL binding to a DR1 sequence recapitulates the known base preferences for binding across the DR1, and captures the previously reported preference for the sequence 5'-ACT-3' upstream of the DR-site (Fig. 3.2a) (Juge-Aubry et al., 1997), demonstrating that my nextPBM experiments accurately capture PPARγ dimer binding specificity.

To examine the PPARγ-AL binding specificity landscape more broadly, I compared the full binding profiles (i.e., across all DR sites) of purified PPARγ:RXRα (Penvose et al., 2019) and PPARγ-AL (Fig. 3.2b). PPARγ-AL showed a lower dynamic range of binding than that of purified PPARγ:RXRα, likely due to a lower concentration of PPARγ in the lysate samples and/or to competition with other NRs for dimerization with RXRα. To avoid any complications arising from the different dynamic ranges, I chose to only consider sequences that were bound by purified PPARγ:RXRα with high-affinity (z-score > 7, Fig. 3.2b blue and light blue) in subsequent analyses, as this is also the inflection point at which PPARγ-AL shows binding above background. At any given affinity for purified PPARγ:RXRα, the binding of PPARγ-AL spans a broad range of affinities. For example, for sites bound by purified PPARγ:RXRα with a z-score ~10 (Fig. 3.2b, red rectangle), PPARγ-AL binds over a range of affinities, from a z-score of -0.3 to 9.1. Strikingly, many sites bound with high affinity by purified PPARγ:RXRα are bound with low affinity by PPARγ-AL (Fig3.2b, light blue dots).

The binding data for purified PPARγ:RXRα dimers was generated for human protein sequences expressed and purified from *E.coli*. The species of the expression host

is known to impact protein folding and post-translational modifications that may affect protein function (Structural Genomics Consortium et al., 2008).To test whether the alteration in binding that I observe in Fig. 3.2b is due to differences between bacterial and mammalian expression systems, I measured the binding of PPARγ:RXRα purified from HEK-293T cells, a mammalian cell line of human origin, and compared it to the binding of PPARγ:RXRα purified from *E. coli*. I found strong agreement in binding of PPARγ:RXRα from both expression hosts (Fig. 3.2c, $R^2 = 0.90$), suggesting that the binding differences observed between purified and lysate samples are not due to the means of protein expression.

To determine if the differences in PPARγ binding observed between purified and lysate samples are a general phenomenon occurring on all DNA sequences, or whether they are specific to certain DR spacer lengths, I compared the binding affinity of PPARγ-AL to at each spacer length. I focused my analysis on sequences bound with high affinity by purified PPARγ:RXRα (z-score > 7, Figure 3.2b, dark and light blue dots). The binding of PPARγ-AL was significantly different across the different spacer lengths (Fig. 3.2d, $p < 0.0001$ by one-way ANOVA on ranks). PPARγ-AL showed a strong preference for high-affinity binding only to DR1 sites (Fig. 3.2d, $p < 0.0001$ by Dunn's Test for all comparisons to DR1). Binding was also significantly different across spacer lengths for purified PPARγ:RXRα, (Fig. 3.2e, $p < 0.0001$ by one-way ANOVA on ranks). In contrast to PPARγ-AL, purified PPARγ:RXRα had a significant preference for binding to both DR1 and DR4 sites (Fig. 3.2e, see figure for individual comparisons). For purified PPARγ:RXRα, the preference for specific spacer lengths were not as strong as the

preference seen in PPARγ-AL (comparing figure 3.2d and 3.2e). DR1 sites are the canonically reported spacer preference for PPARγ (Evans and Mangelsdorf, 2014; IJpenberg et al., 1997; Palmer et al., 1995), and are enriched in ChIP-seq experiments measuring in vivo binding of PPARγ (Nielsen et al., 2008; Siersbaek et al., 2012). This refinement in the DR spacer preferences of PPARγ-AL supports the hypothesis that the DNA binding-site preferences of NRs is altered, and becomes more specific, in the presence of the full complement of soluble nuclear components.

3.3.3 SRC-1 is recruited to DNA in a sequence-specific manner

To test whether the disconnect between NR binding and activity might also be a result of selective recruitment of CoRs to specific DNA sites, I examined how CoR recruitment by NRs correlated with NR binding. We have recently extended the nextPBM methodology to measure recruitment of CoRs to DNA-bound TFs. NextPBM experiments are performed using nuclear extract from cells, which contains both TFs and CoRs. By utilizing an antibody for the CoR of interest, we are able to detect indirect recruitment of CoRs to DNA-bound TFs (schematized in Fig. 3.3a). Furthermore, using our SNV-based approach described for generating NR binding logos, we can determine CoR recruitment logos, which are analogous to TF binding logos but directly reflect binding the specificity of the larger multi-protein TF-CoR complex.

The steroid receptor coactivators (SRCs) are CoRs of NR function, serving as scaffolds to recruit other CoRs and to bridge NRs to the basal transcriptional machinery (Johnson and O'Malley, 2012). SRC1 knockout mice show increased obesity due to a decrease in energy expenditure, positioning SRC1 as a key coregulator of adipocyte

function (Picard et al., 2002). Over the course of adipogenesis in 3T3-L1 adipocytes, SRC1 expression increases dramatically (Picard et al., 2002). SRC1 has been shown to be recruited to many NRs including PPAR, FXR, THR, and RXR (Dowell et al., 1997; Kim et al., 1999; Treuter et al., 1998; Wang et al., 2018b). For these reasons, I chose SRC1 for my pilot experiments of CoR recruitment from 3T3-L1 adipocyte lysate.

I measured the recruitment of SRC1 from 3T3-L1 adipocyte lysate using nextPBM. As SRC1 is recruited by RXRα-containing complexes, I anticipated that SRC1 recruitment profiles and RXRα binding profiles would be highly similar. Surprisingly, I observe distinct DNA binding/recruitment profiles, as can be seen by the lack of correlation between the measured SRC1 and RXRα profiles (Fig. 3.3b). For sites that bind RXRα with high-affinity, SRC1 shows a range of recruitment, from a z-score of -2 up to a z-score of 6, demonstrating that SRC1 is not recruited well by all RXRα−containing complexes (Fig. 3.3b, blue box). In contrast, SRC1 is also recruited with high-affinity to sites that show weak RXRα binding (Fig. 3.3b, red box), suggesting that SRC1 is recruited by TFs other than RXRα-containing complexes. SRC1 is known to be recruited to other NRs, including the estrogen related receptors (ERRs) and hepatocyte nuclear factor 4 (HNF4), which also bind to the DR-sites that are prevalent on this custom-designed PBM (Albers et al., 2005). Further experiments will be needed to determine if SRC1 recruitment is occurring to these other NRs.

As CoRs rely on specific interactions with TFs for recruitment to DNA, logos of CoR recruitment represent the binding preferences of the TF-CoR complex. Our SNV approach gives us the unique advantage of modeling binding for a single DNA seed

sequence, which enables us to observe any changes in TF binding preferences induced by CoRs as alterations in the recruitment logos (eg. allosteric changes in TF binding can be observed as changes in recruitment logos). Comparing the logos obtained for SRC1 to those for RXRα on the same seed sequence, I see distinct DNA binding preferences for these two factors (Fig.3.3c). On all of these seeds, RXRα binds to full DRs, making sequence specific contacts with both half-sites (Fig. 3.3c, bottom row of logos, grey shading indicates half-sites of the DR seed). SRC1 logos reflect general NR binding preferences (i.e., 5'-RGKTCA-3' sites); however, in contrast to the RXRα logos, SRC1 recruitment logos demonstrate that recruitment occurs in many different modes, both to half-sites and full DRs (Fig. 3.3c, top row of logos, grey shading indicates half-sites of the DR seed). The strong similarity of binding preferences seen in the SRC1 recruitment logos and NR binding logos indicates that SRC1 is likely being recruited to DNA by a NR. However, these altered base preferences suggest that either these SRC1-NR complexes do not contain RXRα, or that SRC1 interaction induces an altered DNA-binding specificity to the RXRα-containing complex, constituting a form of allosteric regulation of SRC1 recruitment.

Examining SRC1 logos (Fig. 3.4) in detail, I observe distinct sequence features underlying SRC1 recruitment, extending beyond the core NR-like half-site(s). A strong preference for an A in the position immediately 5' of the NR-like half-site is seen in all SRC1 recruitment logos (Fig. 3.4, red box). For some seed sequences, strong sequence preferences for SRC1 recruitment are also seen in logos at positions more 5' of the NR-like half-site, with a preference for C or T in the position 3 nucleotides upstream of the

NR-like motif and for G or C in the position 2 nucleotides upstream of the NR-like motif (Fig. 3.4c, black box). These recruitment preferences hold true for all logo types (extended 5' half-site, extended 3' half-site, or extended full-site), with the extended binding preferences always occurring 5' of the NR-like half-site. In the case of an extended full-site, these preferences are seen in the 5'-flanking sequence of both half-sites (Fig. 3.4, black and red boxes of extended full DRs, column 3 of logos). These findings suggest that SRC1 is preferentially recruited to DNA sequences with specific sequence features, which may arise from recruitment to specific NRs, recruitment of CoRs only to NRs bound to DNA with a certain sequence composition (eg. allosteric recruitment), or both.

3.4 Discussion

Recent work from our lab demonstrated that purified NRs display a broader DNA binding specificity than had previously been described in the literature (Penvose et al., 2019). Additionally, we demonstrated that models that conform to the canonically reported spacer preferences for NR are more enriched in 'functional binding' events from ChIP-seq data (Penvose et al., 2019). Based on these findings we hypothesized that NR binding may be altered in a cellular context. To test this, I used nextPBM to measure the binding of NRs from 3T3-L1 adipocyte nuclear lysate. To address whether CoR might also be selectively recruited by NRs to specific sequences, in a novel extension of the nextPBM platform, I also measured recruitment of the NR coregulator, SRC1.

I found that the binding preferences of PPARγ from nuclear lysate are altered compared to the binding of purified PPARγ:RXRα. PPARγ from lysate preferentially binds to DR1s, in agreement with both the in vitro and in vivo binding preferences previously reported by many others (Fu et al., 2005; Juge-Aubry et al., 1997); however, the high-affinity binding to DRs of many spacer lengths that we reported for purified PPARγ:RXRα (Penvose et al., 2019) suggests that this preference is not purely determined by DNA binding affinity. The factors that lead to the alteration in binding seen between purified and lysate PPARγ remain unclear. Additionally, the extent to which the binding landscape of other NRs is altered in nuclear lysate requires further investigation.

3T3-L1 cells are a fibroblast cell line derived from mice (Kassotis et al., 2017) The DNA binding preferences of the type II NRs presented in Chapter 2 of this thesis were measured using proteins that were expressed from human coding sequences; however, I do not believe this alteration in binding is likely to be attributed to species-dependent differences in DNA binding. PPARγ is highly conserved in mice and humans, sharing complete amino acid identity within the DBDs and differing by only 9 amino acids total for PPARγ1 (Pap et al., 2016). Additionally, their reported DNA binding preferences are highly similar, (Chandra et al., 2008; Soccio et al., 2015). To experimentally confirm that the altered PPARγ in nuclear lysate from 3T3-L1s is not a species-specific difference in DNA binding preferences, I could perform nextPBM experiments using nuclear lysate from a cell line of human origin. Alternatively, the easiest experiment to interpret would be one in which I co-incubated 3T3-L1 adipocyte

nuclear lysate with His-tagged human PPARγ and RXRα and detected binding of PPARγ using an antibody for the His-tag. If the same alteration in binding preferences was observed in this experiment, it would suggest that some component of the nuclear lysate alters the binding preferences of PPARγ.

An additional complication in interpretation of the alteration in binding preferences of PPARγ -AL is the multiple isoforms of PPARγ present in 3T3-L1 adipocytes. 3T3-L1 adipocytes express both PPARγ1 and PPARγ2 isoforms (Desvergne and Wahli, 1999). The PPARγ2 isoform arises from alternative splicing, and is expressed highly only in adipocytes (Desvergne and Wahli, 1999). PPARγ2 encodes for an additional 28 (human) or 30 (mouse) amino acids at the N-terminus of the protein as compared to the PPARγ1 isoform. The PPARγ antibody used in these experiments recognizes both the PPARγ1 and γ2 isoforms. It is possible that the PPARγ2 isoform of the protein has different DNA binding preferences than the PPARγ1 isoform used in purified protein experiments. The spike-in experiment with His-tagged PPARγ and RXRα proposed above to test species-specific PPARγ binding preferences would also allow us to avoid complications in interpreting the data that may arise from any isoform-specific DNA binding preferences for PPARγ.

I found that SRC1 is preferentially recruited to DNA sequences with a distinct base composition. This finding suggests that either 1) SRC1 is only recruited to certain NRs, or 2) SRC1 is only recruited to NRs bound to DNA with certain sequence features (e.g. allosteric recruitment to DNA). These interpretations are not mutually exclusive,

and both are supported by existing evidence. In support of the first interpretation, SRC1 is known to recruit strongly to only a subset of NRs. Existing evidence supporting the second interpretation includes the allosteric recruitment of coregulators that has been seen for other NRs, including ER, GR, and VDR (Koszewski et al., 2000; Meijsing et al., 2009; Schöne et al., 2016). Further experiments will be required to distinguish these two possible interpretations of preferential SRC1 recruitment.

Few methodologies exist to measure and model CoR recruitment in a high-throughput manner. As TFs rely on interactions with CoRs to mediate their function, using the nextPBM to measure CoR recruitment offers the potential to create and test models of functional TF binding.

3.5 Materials and Methods

3.5.1 3T3-L1 Culture and Differentiation

NIH 3T3-L1 (ATCC: CL-173) pre-adipocytes were maintained in pre-adipocyte media (high-glucose DMEM with 10% calf serum, 100 U/ml penicillin, 100 µg/ml streptomycin, and 0.25 µg/ml amphotericin B). Cells were passaged when they reached ~70-80% confluence, approximately every 3 days. Before inducing adipogenesis, cells were grown as above until they reached confluence, about 4 days from plating. To induce adipogenesis, the medium was replaced with DMEM containing 10% fetal bovine serum (FBS) (Sigma-Aldrich), 250 nM dexamethasone, 167 nM human insulin, 0.5 mM isobutyl methyl xanthine, 100 U/ml penicillin, and 100 µg/ml streptomycin. Cells were treated with either vehicle (dimethyl sulfoxide (DMSO), 0.1% final concentration) or

rosiglitazone (20 μM, Cayman Chemicals). On days 3 and 5 of differentiation, the medium was replaced with adipocyte maintenance medium (DMEM with 10% FBS, 167 nM human insulin, 100 U/ml penicillin, and 100 μg/ml streptomycin), and the cultures were supplemented with either vehicle or rosiglitazone. On Day 7 of differentiation, medium was replaced with adipocyte medium (DMEM, 10% FBS, 100 U/ml penicillin, 100 μg/ml streptomycin), and the cultures were supplemented with either vehicle or rosiglitazone. On day 10 after the initiation of differentiation, nuclear lysates were harvested as described below.

3.5.2 Nuclear extract preparation

Differentiated 3T3-L1 adipocyte cells were scraped from flasks in ice cold PBS. When cells were grown in T75 flasks, three flasks were combined and processed as a single sample using the protocol below. If using T225 flasks, each flask was processed individually as a single sample. All volumes given in the protocol below are for a single sample. Cells were pelleted at 500 x g for 5 min at 4°C in a 15 ml conical tube. The pellet was resuspended and washed with 10 ml of PBS and pelleted at 500 x g for 5 min at 4°C. To remove residual salt remaining from the PBS buffer, the cells were resuspended in 1 ml of low salt buffer (10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl with 1 μl protease inhibitor cocktail (Sigma-Aldrich, cat # P8340) and 1 mM DTT), transferred to a 1.5 ml tube, and pelleted immediately at 500 x g for 5 min at 4°C. The cell pellet was resuspended in 1 ml of low-salt buffer and incubated for 10 min on ice. Fifty μl of 5% IGEPAL (Sigma-Aldrich, cat # I8896) was added to the cell suspension and vortexed for 10 sec. Cells were incubated for an additional 10 min on ice. Nuclei were pelleted at

750 × g for 5 min at 4°C. The supernatant was removed from the nuclear pellet and saved as the cytosolic fraction. To remove residual cytosolic proteins from the surface of the nuclear pellet, 100 µl of the low-salt buffer was gently pipetted onto the side of the tube and allowed to wash the pellet, with care taken not to disrupt the nuclear pellet. This wash was then gently removed and combined with the cytosolic fraction. Next, 200 µl of high-salt buffer (20 mM HEPES (pH 7.9), 25% glycerol, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 420 mM NaCl with 1 µl protease inhibitor cocktail and 1 mM DTT) was pipetted on the pellet and vortexed for 30 s followed by nutation at 4°C for 1 h. The nuclei were pelleted at 4°C for 20 min at 21,000 × g. The supernatant was transferred into a new 1.5 ml tube and labeled as the nuclear soluble protein fraction. Protein concentration of the nuclear lysate was quantified using the Pierce Coomassie Bradford Plus reagent (Pierce) following the manufacturer's microplate protocol. From either 3-T75 flasks or a single T225 flask, I typically obtained lysate at a concentration of ~4-5 mg/ml with a total protein yield of ~ 0.8 -1.0 mg. Nuclear lysate samples were aliquoted into single use samples and snap frozen in liquid nitrogen and stored at -80°C until use.

3.5.3 PBM custom design

PBM experiments were performed using custom-designed microarrays (Agilent Technologies Inc. AMADID 084387, 4 × 180 K format). PBM probes contain a 24 nt constant primer region, a 34 nt variable region, and a 5' GC dinucleotide cap (probe sequences can be found in Supplementary Data 4 of (Penvose et al., 2019)). For each unique SNV probe sequence, five replicate probes were included in each orientation (10 probes per unique sequence). For all other probe sequences four replicate probes were

included with the 34 nt variable region in each orientation (8 probes per unique sequence).

SNV probes: DR seed sequences, defined by two 6-bp half-sites and a variable spacer (0–5 bp), were aligned within the 34 nt variable region of each PBM probe. For each seed sequence, SNV probes were created that had a single-nucleotide variant at each position of the DR half-sites, the spacer sequence between the DR half-sites, and in the 5 bp flanks of each site. Therefore, for a single 13 bp DR1 site (i.e., 6 + 6 + 1 = 13), including 5 bp flanks on either side, there would be 69 (i.e., 23 × 3) unique SNV probe sequences.

Random genomic probes: 34 nt regions were randomly chosen from the UCSC hg19 build of human genome. Sequences were removed that contained Ns or single-nucleotide repeats longer than three nucleotides.

3.5.4 NextPBM Experiments

Double stranding – Microarrays were double-stranded as previously described (PBM double-stranding primer 5'-CCTTCATTCTACGCTGTCAATCGC-3') (Berger and Bulyk, 2009; Berger et al., 2006b).

Blocking – All wash steps were carried out in coplin jars on an orbital shaker at 125 rpm. Double-stranded DNA microarrays were first pre-washed in PBS containing 0.01% Triton X-100 (5 min), rinsed in a PBS bath, and then blocked with 2% milk in PBS for 1 hour. Following the blocking step, arrays were washed in PBS containing 0.1%

Tween-20 (5 min), then in PBS containing 0.01% Triton X-100 (2 min), and finally briefly rinsed in a PBS bath.

Protein binding – Arrays were next incubated with the protein sample (purified protein or 3T3-L1 nuclear lysate) for one hour in a binding reaction buffer containing: 2% milk (final concentration); PBS pH 7.4; 1 mM DTT; 0.2 mg/ml BSA; 0.02% Triton X-100; and 0.4 mg/ml salmon testes DNA (Sigma D7656). For nuclear lysate samples, the final concentration of total protein in the protein binding mixture was 2.0 mg/ml in a volume of 180 μL. Concentrations of purified proteins used for PBM experiments can be found in Table 2.1.

Primary antibody – Microarray sandwiches were disassembled in PBS containing 0.05% Tween-20 (3 min) and then de-wetted in plain PBS. Microarrays were then incubated with 20 μg/ml of primary antibody in 180 μL of 2% milk in PBS for 20 min. Primary antibodies used were anti-PPARγ antibody (Abcam ab41928), anti-RXRα (Active Motif 61029), and anti-SRC1 (Abcam ab2859).

Secondary antibody - Microarray sandwiches were disassembled in PBS containing 0.05% Tween-20 for 3 min and then de-wetted in plain PBS. Microarrays were then incubated with 20 μg/ml of Alexa488-conjugated secondary antibody or Alexa647-conjugated secondary antibody in 180 μL of 2% milk in PBS for 20 min. Excess antibody was removed by washing with PBS containing 0.05% Tween-20 for 3 min and then PBS for 2 min.

PBM data analysis - Microarrays were scanned with a GenePix 4400A scanner and fluorescence was quantified using GenePix Pro 7.2. Exported data were normalized using MicroArray LINEar Regression (Berger et al., 2006b). Microarray probe sequences and fluorescence values from each experiment are provided (See Supplementary Data 4 of (Penvose et al., 2019)). NR dimers exhibit an orientation-specific bias in our PBM experiments; therefore, data from probes in a single orientation (i.e., '_o1′ probes in Supplementary Data 4 of (Penvose et al., 2019)) was used in our final analysis. However, all results were observed for probes in both orientations and models from each orientation showed good agreement.

DNA-binding logos were generated using the previously described SNV-based approach (Andrilenas et al., 2018; Penvose et al., 2019).

3.6 Limitations of the nextPBM platform

3.6.1 Competition for RXR as a partner

In both in vitro and in vivo assays when multiple NRs are present in a mixture, NRs compete for dimerization with RXR as a partner (Miyata et al., 1996, Tobin et al., 2000, Chan and Wells, 2009). Decreased binding of LXR/RXR is observed in EMSAs when PPAR is added to the binding reaction (Tobin et al., 2000). In cells, overexpression of PPAR leads to a dose-dependent decrease in expression of an LXR-regulated reporter gene expression (Tobin et al., 2000). In cells that are overexpressing PPAR, co-immunoprecipitation of RXR with LXR is reduced as compared to cells in which PPAR is not overexpressed (Tobin et al., 2000). In all cases, these effects are not seen when

supplemental RXR is provided, suggesting that limiting quantities of RXR underlie these observations, rather than direct competition for DNA binding (Tobin et al., 2000).

Considering these findings, it is likely in the mixture of proteins from nuclear lysate used on nextPBM, we are observing competition amongst NRs for RXR as a dimeric partner; however, this competition does not explain why the decrease in DNA binding of PPARγ from lysate is DR spacer-length dependent. In other words, if the decreased binding of PPARγ in lysate is a consequence of the effective concentration of the PPARγ:RXR heterodimer due to competition for RXR as a partner, one would expect that sites that are bound with equal affinity by the purified heterodimer would show a uniform decrease of PPARγ:RXR binding from lysate, which is not what I observed. Therefore, while competition for RXR as a partner is a factor that must be considered when interpreting nextPBM results, other factors are likely contributing to the altered binding preferences of PPARγ.

3.6.2 Competition for DNA binding sites

An alternative explanation for the altered DNA binding preferences of PPARγ from lysate observed on nextPBM could be that multiple NRs in the lysate are competing for binding on certain DNA sequences. Competition for binding to DRs has been reported by other groups (Chan and Wells, 2009). However, I do not think that the PPARγ-AL binding preferences I found by nextPBM are explained by competition for DNA binding sites. Our lab has tried to directly measure competition on the PBM using purified proteins at concentrations higher than those of the NRs in lysate and we were unable to

detect competition (e.g., the binding profiles obtained for the individual NRs were the same as those when the NRs were co-incubated on the array). This lack of competition for DNA binding sites is likely explained by the high concentration of DNA on the PBM, though the exact amount of DNA at any given array spot is proprietary information that the manufacturer (Agilent) does not publicly disclose.

3.6.3 Multiple NR isoforms and isotypes present in cells

NRs exist as different isoforms, and multiple isoforms are expressed in a single cell-type, such as the 3T3-L1 adipocytes for the experiments in this chapter. For instance, as discussed in Section 3.4, PPARγ is present in 3T3-L1 adipocytes in both the γ1 and γ2 isoform. It is possible that different isoforms have distinct DNA binding preferences. Additionally, humans and mice each contain three paralogous genes for RXR (α, β, and γ) (Desvergne and Wahli, 1999). In NR biology, the paralogs of individual NRs are commonly referred to as isotypes (Desvergne and Wahli, 1999). PPARγ is able to dimerize with any of the three RXR isotypes. RXR isotypes serve partially redundant functions, but are not entirely able to compensate for one another. For instance, adipocyte-selective knock-out of RXRα leads to impaired induction of lipogenic gene programs and resistance to obesity in mice; however, this effect is not seen in RXRγ-null mice (Metzger et al., 2005). Additionally, adipocyte formation is not impaired for adipocyte-selective knock-out of RXRα nor in RXRγ-null mice, but is impaired in adipocyte-selective knock-out of RXRα in RXRγ-null mice (Metzger et al., 2005). The presence of similar isoforms and isotypes of NRs in cells complicates the interpretation

of nextPBM binding data as different isoform and isotype dimers may have distinct binding preferences, and the current approaches used to measure binding do not delineate between distinct isoform and isotype complexes (e.g., when individually measuring the binding of PPARγ, we can look at the correlation of binding of RXRα at the same probes to infer the heterodimeric partner, but it is possible that PPARγ forms heterodimers with all the RXR isotypes. This makes it difficult to attribute the binding observed on nextPBM experiments to a single PPARγ:RXR heterodimer, hence making it more difficult to directly compare the binding preferences to a single purified protein complex. Despite this, nextPBMs still represents a way for us to capture the binding preferences of TF complexes in a context that more closely resembles a cellular context.

3.6.4 Antibody selection for nextPBM experiments

One of the greatest technical challenges with the nextPBM is finding antibodies that perform well in nextPBM experiments. Antibodies are considered "good" on nextPBM when experiments show a broad dynamic range of signal and the models generated show sequence-specificity (e.g., the models resemble existing models or are reproducible). As a starting point, whenever possible, we use antibodies that are ChIP-grade, as these antibodies have been shown to recognize native protein when bound to DNA and to selectively recognize their target protein within a complex mixture of proteins. In cases where no ChIP-grade antibodies are available, we choose antibodies that have been demonstrated to work in assays where the protein is in a native conformation, such as EMSA supershift or immunoprecipitation experiments. Even with this initial selection step, we still find that many antibodies do not perform well on

nextPBM, either giving low overall signal or showing poor specificity (e.g., high background). Failure of antibodies to perform well in nextPBM experiments may be due to a variety of factors including epitope masking, lack of expression of the protein of interest in the lysate, background binding to DNA alone, and/or binding to non-target proteins.

Even for antibodies that perform well on the nextPBM, a variety of factors could influence interactions of the antibody with its target protein. This may be a more pronounced problem in nextPBM experiments as compared to PBM experiments using purified proteins, as masking of the epitope recognized by the antibody is more likely to occur in lysate due to factors such as the potential PTMs present on the protein and interactions with coregulators that occlude portions of the TF. It is also possible that the epitope that the antibody recognizes may be inaccessible due to structural changes of the TF when binding certain DNA probes; or that DNA-induced structural changes in the protein may change the affinity of the protein-antibody interaction on some probes. In the case of CoRs, certain antibodies may disrupt TF-CoR interactions. These complications highlight the importance of supporting any findings on nextPBM using orthogonal methodologies.

Despite the complications presented here, the nextPBM methodology offers a novel way to measure and model protein-DNA and coregulator-protein-DNA interactions in an environment that better simulates a cellular context.

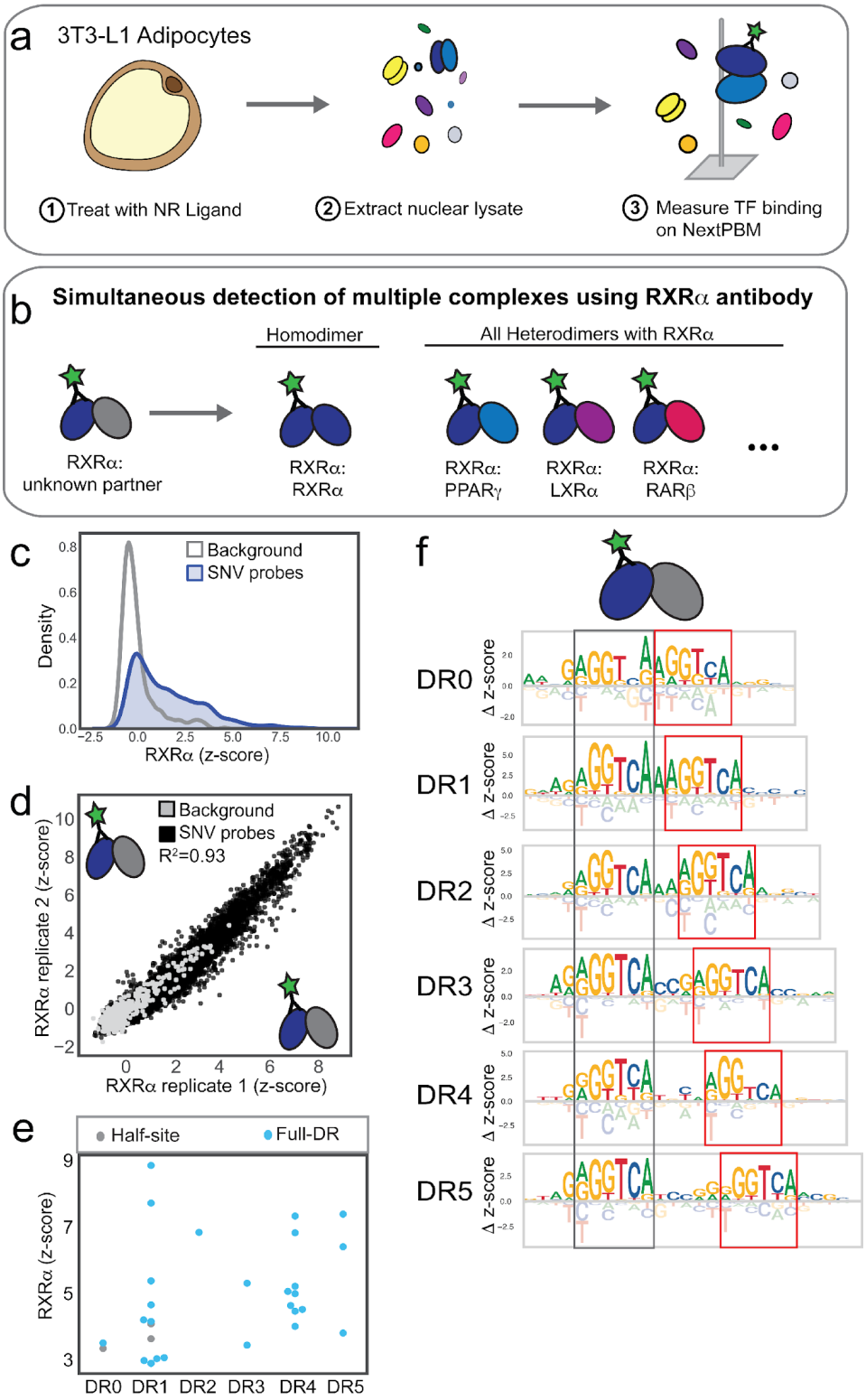


Figure 3.1: NR binding from 3T3-L1 adipocyte nuclear lysate examined by nextPBM

a Schematic overview of NextPBM experiment using 3T3-L1 adipocyte nuclear lysate. **b** Schematic demonstrating that a single RXRα antibody allows assessment of binding for all RXRα–containing complexes. **c** PBM replicate averaged z-score distributions for RXRα-containing complexes from 3T3-L1 adipocyte nuclear lysate for all SNV probes (blue) and genomic background probes (grey). **d** Scatter plot of z-scores of binding for two independent biological replicates for RXRα-containing complexes from 3T3-L1 adipocyte nuclear lysate. Dots represent average over ~5 replicates for all 10,728 unique SNV probes (black dots) and 500 background probes (gray dots). **e** At each DR spacer length, the replicate averaged z-score for each seed sequence is shown; seed sequences with z-score < 3 are not represented. Colors indicate model-type for each seed sequence. **f** Representative logos from seeds at each DR spacer length. For each logo, the 5' half-site is boxed in grey and the 3' half-site is boxed in red.

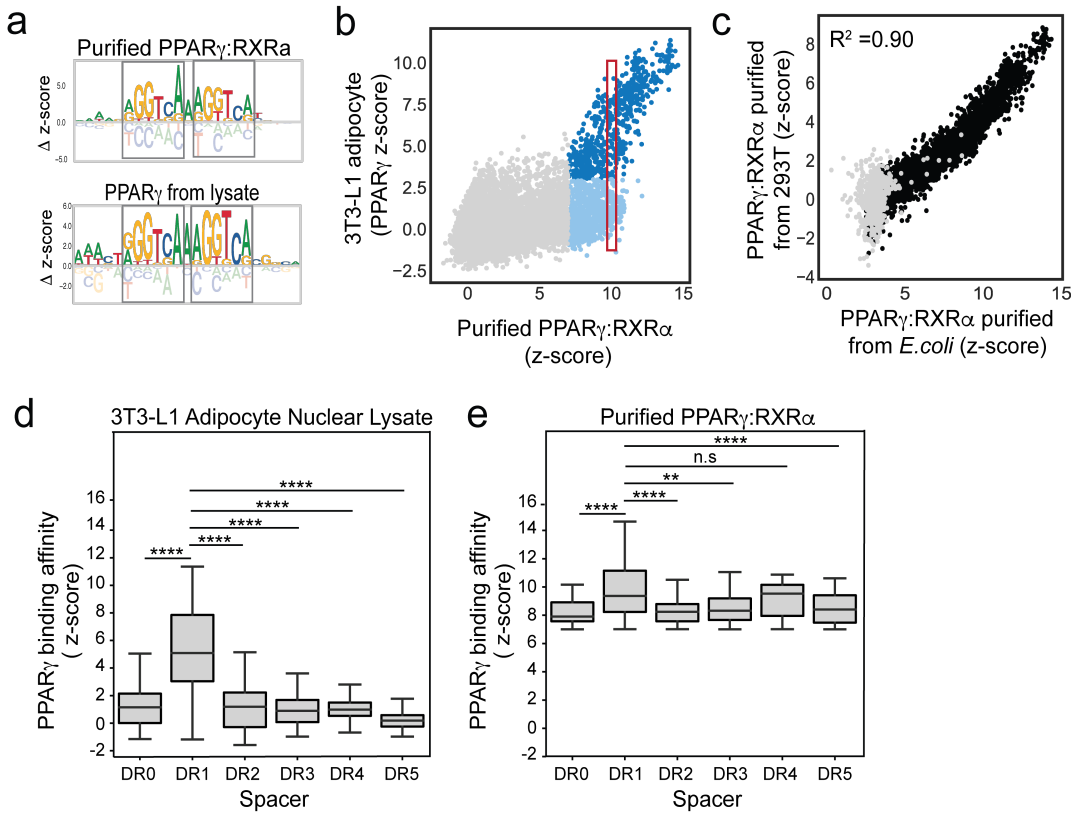


Figure 3.2: PPARγ binding landscape from 3T3-L1 adipocyte nuclear lysate

a Logos for purified PPARγ:RXRα and PPARγ from 3T3-L1 adipocyte nuclear lysate bound to a DR1 seed. Grey boxes indicate each half-site of the direct repeat within the seed. **b** Scatter plot comparing z-scores for PPARγ from 3T3-L1 adipocyte nuclear lysate to purified PPARγ:RXRα on all SNV probes. Colors represent classification of probes as high-affinity (dark and light blue dots) or low-affinity (grey dots) for purified PPARγ:RXRα. **c** Scatter plot comparing z-scores of binding of PPARγ:RXRα purified from 293T cells and PPARγ:RXRα purified from *E. coli*. Dots represent average over ~5 replicates for all 10,728 unique SNV probes (black dots) and 500 background probes (gray dots). **d, e** At each DR spacer length, the distributions of z-scores of binding for PPARγ from 3T3-L1 adipocyte nuclear lysate (3.2d) or purified PPARγ:RXRα (3.2e) are shown. Data are shown only for probes bound by purified PPARγ:RXRα with a z-score >7.0. Center line: median; box limits: upper and lower quartiles; whiskers: last datum within 1.5x interquartile range. Differences in binding to certain spacer lengths were tested by one-way ANOVA by ranks followed individual comparisons using Dunn's test with correction for multiple hypothesis testing. n.s denotes no significant difference, ** denotes $p < 0.01$, **** denotes $p < 0.0001$ as calculated by Dunn's test with correction for multiple hypothesis testing.

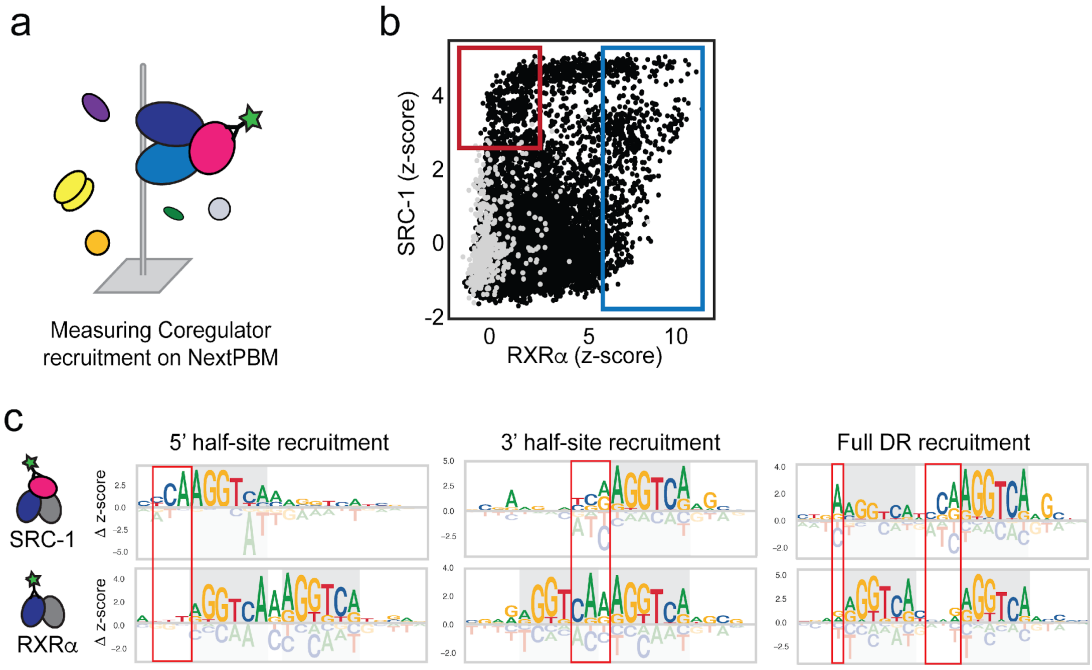


Figure 3.3: Measuring SRC1 recruitment from 3T3-L1 adipocyte nuclear lysate.

a Schematic overview of NextPBM experiment using 3T3-L1 adipocyte nuclear lysate to measure coregulator recruitment. **b** Scatter plot of z-scores of binding of SRC1 and RXRα-containing complexes from 3T3-L1 adipocyte nuclear lysate. Dots represent average over ~5 replicates for all 10,728 unique SNV probes (black dots) and 500 background probes (gray dots). Red box indicates high-affinity SRC1 binding and low affinity RXRα binding. Blue box indicates the range of SRC1 recruitment occurring to sites that RXRα binds with high-affinity. **c** Representative logos of RXRα binding and SRC1 recruitment determined using the same seed and SNV sequences. Grey highlights indicate half-sites within each probe sequence. Red boxes indicate flank sequences of SRC1 recruitment that show distinct base preferences compared to RXRα.

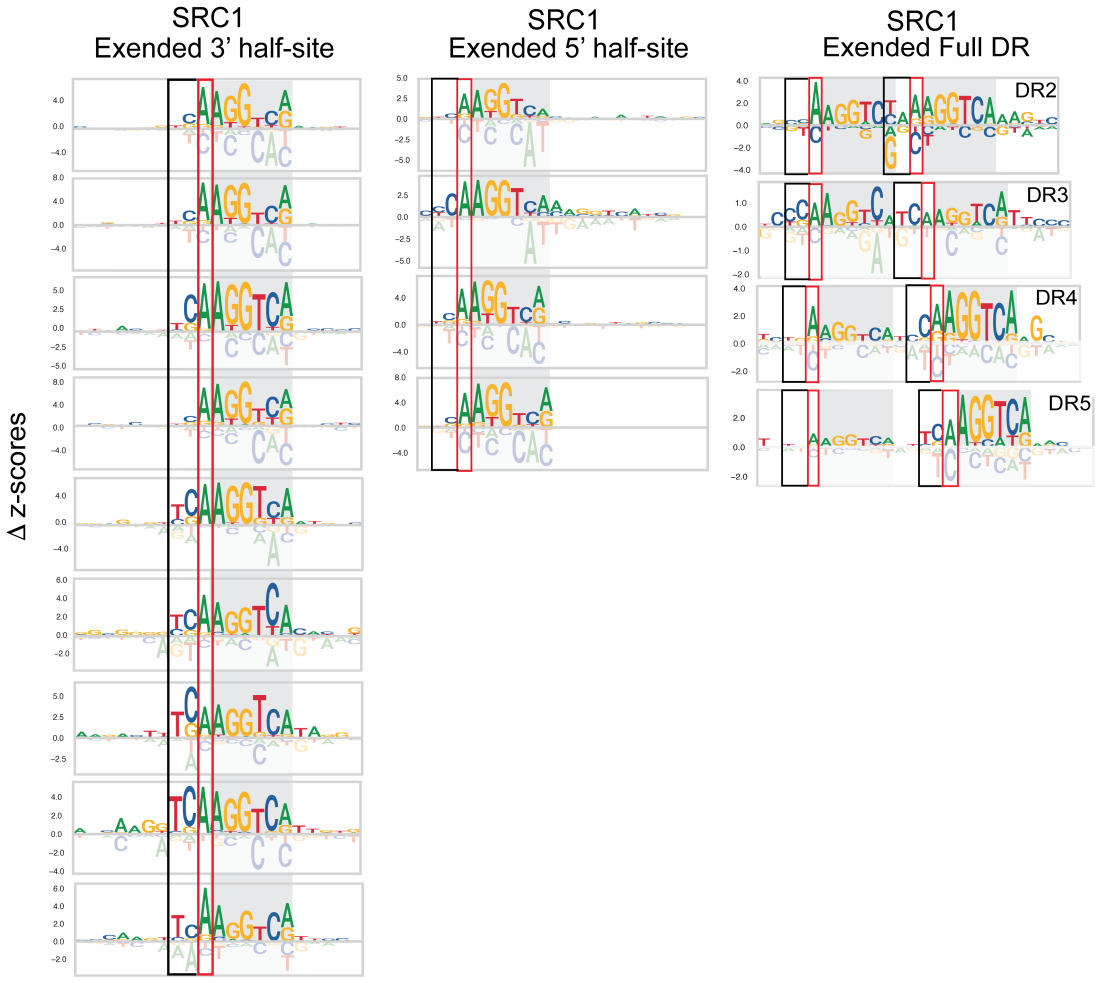


Figure 3.4: SRC1 shows distinct DNA sequence preferences for recruitment

Logos of SRC1 recruitment on different seed sequences. Grey highlight indicates the half-sites within each probe sequence. Red and black boxes indicate flank sequences of SRC1 recruitment. DR(2-5) indicates the design of the seed probe.

Antibody	Source	Species	Epitope	NextPBM notes
LXRβ	Abcam ab56237	Rabbit	Synthetic peptide corresponding to Human LXRβ aa 37-86	Medium signal and good specificity
PPARγ	Abcam ab41928	Mouse	Baculovirus-expressed recombinant fragment, corresponding to amino acids 3-105 of Human PPARγ 1+2.	Good signal and specificity
RXRα	Active Motif 61029	Mouse	Recombinant protein corresponding to C-terminal region of RXRα	Good signal and specificity
RIP140	Abcam ab42126	Rabbit	Recombinant protein corresponding to C-terminal region of RIP140	Low overall signal
SRC-1	Abcam ab2859	Rabbit	Synthetic peptide corresponding to Human KAT13A/ SRC1 aa 8-24.	Good signal and specificity
SIRT1	Millipore 07-131	Rabbit	GST-tagged fusion protein corresponding to amino acids 1-131 of mouse SIRT1	Low overall signal
TBL1X	Santa Cruz sc-365661X	Mouse	specific for an epitope mapping between amino acids 119-148 within an internal region of TBL1X of human origin	Low overall signal

Table 3.1: Antibodies tested in nextPBM experiments using 3T3-L1 adipocyte nuclear lysate

Shown are the antibodies that I have tested in nextPBM experiments using 3T3-L1 adipocyte nuclear lysate. The source and species host of the antibody are listed. The manufacturer's notes on the epitope within the protein that is recognized by the antibody is shown. The nextPBM notes column indicates the performance of the antibody in nextPBM experiments.

CHAPTER FOUR: DISCUSSION/ FUTURE DIRECTIONS

4.1 Summary and future directions

The goal of the research presented in this thesis was to explore the determinants of regulatory specificity for the type II NRs. At the time we began this work, existing PWM models of NR binding were not able to uniquely discriminate the genomic targets of NRs. Additionally, DNA had been shown to be an allosteric regulator of CoR recruitment; however, few studies had examined this phenomenon using a high-throughput approach, and this phenomenon had only been studied for a limited set of NRs. In this research, I used PBMs to measure the DNA binding preferences of 12 purified type II NRs. In a follow-up to this work with purified NRs, I used nextPBM to measure binding of NRs from nuclear lysate, and to measure recruitment of SRC1 in the presence of DNA.

In Chapter 2 of this thesis, I demonstrated that the binding preferences of purified NRs do not conform to their previously reported DR-spacer preferences. Discrepancies between our findings and previously reported preferences may result from the different approaches used to establish these preferences. Previous studies that broadly examined the DNA binding preferences of NRs used EMSAs to test DNA binding affinity. The test sequences used were either semi-random oligonucleotide pools, coupled to multiple rounds of PCR to enrich the DNA sequences to which the NR was bound (Kurokawa et al., 1993; Näär et al., 1991), which leads to enrichment for only the highest affinity binding sites and does not capture the full range of NR binding; or were derived from ligand-responsive genes (IJpenberg et al., 1997; Palmer et al., 1995) for the NR in

question, thus biasing the results toward functional NR binding. Additionally, binding affinity in EMSA is estimated from the amount of DNA probe that is shifted in the presence of the TF of interest, which requires that the TF-DNA complex remains assembled throughout the electrophoretic process, and as such may not accurately capture equilibrium binding (e.g., sequences for which TFs show comparable affinity (K_d) but different k_{on} and k_{off} rates may not result in the same amount of shifted probe). An advantage to our PBM approach was the ability to directly compare NR DNA binding preferences on thousands of different DNA sequences at multiple different spacer lengths (DR0-5). Additionally, our SNV modeling approach enables us to create models for individual DNA sequences, allowing for a greater understanding of NR-DNA interactions than was previously possible.

A novel finding of Chapter 2 was the widespread high-affinity binding of NRs to half-site sequences. Half-site binding has been reported for other TFs, including GR and NF-κB (Siggers et al., 2011; Weikum et al., 2017). GR binding in the promoters for pro-inflammatory genes had long been reported to occur through indirect recruitment to another TF, AP-1; however, a recent study demonstrated that recruitment to these sites is not dependent on AP-1, but rather is mediated through direct GR-DNA interactions that occur through binding to GR half-sites that are imbedded within AP-1 binding sites. Similarly, homodimers of c-Rel, a member of the NF-κB TF family, bind with high affinity to NF-κB half-sites to preferentially activate distinct gene targets (Siggers et al., 2011). These findings demonstrate that half-site binding occurs for other TFs that normally function as dimers and highlights the importance of these half-sites in gene

regulation. Further work will be required to delineate the functional importance of half-site binding for the type II NRs.

In Chapter 2, we also demonstrated that canonical DR spacer preferences for an individual NR are enriched in 'functional' genomic binding data for that NR, and correlate with function of that NR as measured in reporter assays. The strong correlation of canonical spacing preferences with NR function may have reinforced the binding preferences for NRs that had been established by EMSA. The findings from Chapter 2 suggest that in cells (1) NR binding preferences are altered by some unknown mechanism(s) and/or (2) NRs can bind more broadly than these canonical rules would suggest, but they mediate transcription only from DRs of certain spacer lengths.

This disconnect between TF binding affinity and TF function is seen for many other transcription factors (Andrilenas et al., 2018). Investigating the mechanisms underlying this disconnect remains difficult, as gene regulation is a multi-step process that is regulated at many levels. Furthermore, determining how these mechanisms, which may involve allostery, may differ between reporter assays and in vivo genomic binding is very complicated. However, we think that these types of integrative comparisons of TF DNA-binding affinity data and reporter gene assay data provide an initial starting point for understanding the relationship between TF binding and TF function that will highlight how pervasive these types of phenomena may be for different classes of TFs.

In Chapter 3 of this thesis, I tested whether binding of NRs is altered in a cellular context by utilizing nextPBM to measure the binding preferences of NRs from 3T3-L1

adipocyte nuclear lysate. I observed high-affinity binding of RXRα-containing complexes to all DR spacer lengths examined (DR0-5). Strikingly, the binding of PPARγ from nuclear lysate showed altered binding preferences, with a strong preference for binding to the canonically reported DR1 site (Palmer et al., 1995). This finding is of particular interest as it raises questions about what mechanisms in a cellular context contribute to these changes in binding. Future nextPBM experiments should seek to determine if the binding preferences of other NRs are also altered in cell lysate. In nuclear lysate, interactions with coregulators may alter NR structure, destabilizing NRs on the "wrong" DR spacer, to refine their binding landscape. To test whether the binding landscape of NRs is altered by interactions with CoRs, the binding of purified NRs in the presence of CoRs could be measured by PBM. As many CoRs are quite large and the domains utilized by different TFs to interact with CoRs is often known, isolated domains of these proteins may make such experiments more tractable.

Promoter-specific TF function has been demonstrated for a variety of TFs, including the NRs (Johnson and O'Malley, 2012); however, the mechanisms dictating such locus-specific function largely remain unclear. Allosteric coregulator recruitment, wherein DNA sequence impacts productive interactions with CoRs without impacting the DNA binding affinity of the recruiting TF, is one mechanism that could explain promoter-specific NR function. Allosteric regulation of coregulator recruitment provides a mechanism for decoupling TF DNA binding affinity from transcriptional activity, as sites not conforming to these preferences may show decreased CoR recruitment and decreased transcriptional upregulation. In section 3.3.3, I used nextPBMs to explore the

possibility that SRC1, a NR CoR, assembles only on a subset of NRs bound to DNA and that certain DNA sequence features can be associated with this recruitment. I found that SRC1 is differentially recruited to RXRα−containing complexes, with SRC1 being recruited strongly to some RXRα−containing complexes and not at all to others. Additional experiments will be required to determine if this finding represents allosteric recruitment of SRC1, binding of SRC1 to only a subset of NRs, or both.

The ability of DNA to act as an allosteric regulator of TF function has been demonstrated by other groups for NF-κB and the glucocorticoid receptor (Meijsing et al., 2009; Wang et al., 2012). Additionally, previous studies have demonstrated that allosteric recruitment of CoRs influences NR function (Meijsing et al., 2009). Genomic analyses using models of SRC1 recruitment generated from nextPBM experiments will allow us test whether SRC1 recruitment preferences better predict functional NR binding in the genome as compared to models of NR binding alone (similar to the analyses performed in Section 2.3.8, however, models of SRC1 recruitment would be utilized rather than models of NR binding).

Additionally, DNA-based allosteric alteration of TF interactions with CoRs and subsequent impact on TF function may provide a mechanism to explain the impact of some single-nucleotide polymorphisms (SNPs) on gene regulation. Interest in precision medicine continues to increase as the decrease in cost of whole-genome sequencing makes it more feasible (Nielsen, 2017). The recent availability of large amounts of genomic data has led the identification of many disease-associated SNPs (Hindorff et al., 2009). The majority of SNPs (>90%) do not fall within the sequence of genes, but rather

in noncoding regions of the genome that are often regulatory regions, such as promoters and enhancers; and the impact of these SNPs is not fully understood (Bryzgalov et al., 2013; Hindorff et al., 2009). Regulatory SNPs have also been shown to correlate with responsiveness to treatment with rosiglitazone, a PPARγ ligand used to treat type II diabetes (Soccio et al., 2015). It is possible that for at least a subset of regulatory SNPs, the SNPs do not alter TF binding, but rather alter CoR recruitment, either leading to decreased or increased interactions with CoR that alter TF function at a SNP. The nextPBM provides a novel high-throughput platform to screen for both changes in TF binding and DNA-based allostery in the context of disease-associated SNPs.

Utilizing the nextPBM platform, we are able to investigate many outstanding questions about how TFs function in a cellular context and interact with other proteins that may be important for their function. TF binding may be altered in a cellular context, due to signal-dependent PTMs or through cooperatively binding to DNA with other TFs; these types of changes can be measured in the nuclear lysate samples used on the nextPBM. Additionally, a large outstanding problem for understanding transcriptional regulation mediated by TFs is the ability to discriminate non-functional TF binding from functional TF binding. The majority of TF binding events measured by ChIP-seq appear not to serve a functional purpose (Spivakov, 2014); and the current models used to predict TF binding (such as PWMs) are unable to determine which binding events will have functional consequences (Evans and Mangelsdorf, 2014). The nextPBM provides a novel platform to measure and model TF-DNA interactions in an environment that more

closely mimics a nuclear context, and to measure and model the ability of TFs to form higher-order transcriptional complexes on specific DNA sequences.

Our lab is also planning to use a modified form of massively parallel reporter assays (MPRA) to test and model TF function. In this technique, synthetic promoter sequences containing the test DNA binding sites of interest are coupled to a reporter gene and unique molecular identifiers (UMIs). These UMIs allow us to map the resulting transcripts back to the test DNA binding sites. Utilizing this modified MPRA, we can test the function of hundreds of DNA sequences (or more) in parallel. The test sequences in this modified MPRA will be designed from models of TF function derived from nextPBM experiments. This assay can be done in the presence or absence of overexpression of the TF of interest, to better enable us to attribute reporter gene expression to a specific TF. Modified MPRAs allow us to test models of functional TF binding generated by nextPBM in a functional assay.

The goal of this thesis was to undertake research into the determinants of type II NR DNA-binding and gene regulatory specificity. This research has resulted in the generation of dozens more NR PWM models than were previously available, all of which can be used to predict NR binding in the genome. The promiscuous DNA binding preferences of the type II NRs and disconnect between NR DNA binding and NR gene regulatory function reveal the need for further investigation into the determinants of functional NR-DNA binding. The findings presented in chapter 3 suggest that DNA binding specificity of NRs are altered in nuclear lysate by some unknown mechanism and that DNA-based allostery may play a role in the recruitment of higher order complexes to

NRs. Further research should seek to clarify the determinants of functional NR binding, changes in NR DNA binding preferences in cells, and the molecular mechanisms underlying both of these phenomena.

BIBLIOGRAPHY

Albers, M., Kranz, H., Kober, I., Kaiser, C., Klink, M., Suckow, J., Kern, R., and Koegl, M. (2005). Automated yeast two-hybrid screening for nuclear receptor-interacting proteins. Mol. Cell Proteomics *4*, 205–213.

Andrilenas, K.K., Penvose, A., and Siggers, T. (2015). Using protein-binding microarrays to study transcription factor specificity: homologs, isoforms and complexes. Briefings in Functional Genomics *14*, 17–29.

Andrilenas, K.K., Ramlall, V., Kurland, J., Leung, B., Harbaugh, A.G., and Siggers, T. (2018). DNA-binding landscape of IRF3, IRF5 and IRF7 dimers: implications for dimer-specific gene regulation. Nucleic Acids Research *46*, 2509–2520.

Balaguer, P., Delfosse, V., Grimaldi, M., and Bourguet, W. (2017). Structural and functional evidences for the interactions between nuclear hormone receptors and endocrine disruptors at low doses. Comptes Rendus Biologies *340*, 414–420.

Barak, Y., Nelson, M.C., Ong, E.S., Jones, Y.Z., Ruiz-Lozano, P., Chien, K.R., Koder, A., and Evans, R.M. (1999). PPAR gamma is required for placental, cardiac, and adipose tissue development. Molecular Cell *4*, 585–595.

Barish, G.D., and Evans, R.M. (2004). PPARs and LXRs: atherosclerosis goes nuclear. Trends in Endocrinology & Metabolism *15*, 158–165.

Berger, M.F., and Bulyk, M.L. (2009). Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. Nature Protocols *4*, 393–411.

Berger, M.F., Philippakis, A.A., Qureshi, A.M., He, F.S., Estep, P.W., and Bulyk, M.L. (2006). Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. Nature Biotechnology *24*, 1429–1435.

Bernardes, A., Batista, F.A.H., de Oliveira Neto, M., Figueira, A.C.M., Webb, P., Saidemberg, D., Palma, M.S., and Polikarpov, I. (2012). Low-Resolution Molecular Models Reveal the Oligomeric State of the PPAR and the Conformational Organization of Its Domains in Solution. PLoS ONE *7*, e31852.

Boergesen, M., Pedersen, T.Å., Gross, B., van Heeringen, S.J., Hagenbeek, D., Bindesbøll, C., Caron, S., Lalloyer, F., Steffensen, K.R., Nebb, H.I., et al. (2012). Genome-Wide Profiling of Liver X Receptor, Retinoid X Receptor, and Peroxisome Proliferator-Activated Receptor α in Mouse Liver Reveals Extensive Sharing of Binding Sites. Molecular and Cellular Biology *32*, 852–867.

Bolotin, E., Liao, H., Ta, T.C., Yang, C., Hwang-Verslues, W., Evans, J.R., Jiang, T., and Sladek, F.M. (2009). Integrated approach for the identification of human hepatocyte nuclear factor 4α target genes using protein binding microarrays. Hepatology *51*, 642–653.

Bryzgalov, L.O., Antontseva, E.V., Matveeva, M.Y., Shilov, A.G., Kashina, E.V., Mordvinov, V.A., and Merkulova, T.I. (2013). Detection of Regulatory SNPs in Human Genome Using ChIP-seq ENCODE Data. PLoS ONE *8*, e78833.

Calkin, A.C., and Tontonoz, P. (2012). Transcriptional integration of metabolism by the nuclear sterol-activated receptors LXR and FXR. Nature Reviews Molecular and Cell Biology *13*, 213–224.

Castrillo, A., and Tontonoz, P. (2004). Nuclear Receptors in Macrophage Biology: At the Crossroads of Lipid Metabolism and Inflammation. Annual Review of Cell and Developmental Biology *20*, 455–480.

Chan, L.S.A., and Wells, R.A. (2009). Cross-Talk between PPARs and the Partners of RXR: A Molecular Perspective. PPAR Research *2009*, 925309.

Chandra, V., Huang, P., Hamuro, Y., Raghuram, S., Wang, Y., Burris, T.P., and Rastinejad, F. (2008). Structure of the intact PPAR-γ–RXR-α nuclear receptor complex on DNA. Nature *456*, 350–356.

Chatagnon, A., Veber, P., Morin, V., Bedo, J., Triqueneaux, G., Sémon, M., Laudet, V., d'Alché-Buc, F., and Benoit, G. (2015). RAR/RXR binding dynamics distinguish pluripotency from differentiation associated cis-regulatory elements. Nucleic Acids Research *43*, 4833–4854.

Choi, J.H., Banks, A.S., Estall, J.L., Kajimura, S., Boström, P., Laznik, D., Ruas, J.L., Chalmers, M.J., Kamenecka, T.M., Blüher, M., et al. (2010). Anti-diabetic drugs inhibit obesity-linked phosphorylation of PPARγ by Cdk5. Nature *466*, 451–456.

Choi, J.H., Banks, A.S., Kamenecka, T.M., Busby, S.A., Chalmers, M.J., Kumar, N., Kuruvilla, D.S., Shin, Y., He, Y., Bruning, J.B., et al. (2011). Antidiabetic actions of a non-agonist PPARγ ligand blocking Cdk5-mediated phosphorylation. Nature *477*, 477–481.

Choudhary, M., and Malek, G. (2016). Rethinking Nuclear Receptors as Potential Therapeutic Targets for Retinal Diseases:. Journal of Biomolecular Screening *21*, 1007–1018.

Chrisman, I.M., Nemetchek, M.D., de Vera, I.M.S., Shang, J., Heidari, Z., Long, Y., Reyes-Caballero, H., Galindo-Murillo, R., Cheatham, T.E., Blayo, A.-L., et al. (2018).

Defining a conformational ensemble that directs activation of PPARγ. Nature Communications *9*, 1794.

Claessens, F., and Gewirth, D.T. (2004). DNA recognition by nuclear receptors. Essays in Biochemistry *40*, 59–72.

Consortium, T.E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Cotnoir-White, D., Laperrière, D., and Mader, S. (2011). Evolution of the repertoire of nuclear receptor binding sites in genomes. Molecular and Cellular Endocrinology *334*, 76–82.

Dasgupta, S., Lonard, D.M., and O'Malley, B.W. (2014). Nuclear Receptor Coactivators: Master Regulators of Human Health and Disease. Annual Review of Medicine *65*, 279–292.

de Aguiar Vallim, T.Q., Tarling, E.J., and Edwards, P.A. (2013). Pleiotropic roles of bile acids in metabolism. Cell Metabolism *17*, 657–669.

Delfosse, V., Grimaldi, M., Cavaillès, V., Balaguer, P., and Bourguet, W. (2014). Structural and Functional Profiling of Environmental Ligands for Estrogen Receptors. Environmental Health Perspectives *122*, 1306–1313.

Desvergne, B., and Wahli, W. (1999). Peroxisome proliferator-activated receptors: nuclear control of metabolism. Endocrine Reviews *20*, 649–688.

Dowell, P., Ishmael, J.E., Avram, D., Peterson, V.J., Nevrivy, D.J., and Leid, M. (1997). p300 functions as a coactivator for the peroxisome proliferator-activated receptor alpha. Joural of Biological Chememistry *272*, 33435–33443.

Evans, R.M., and Mangelsdorf, D.J. (2014). Nuclear Receptors, RXR, and the Big Bang. Cell *157*, 255–266.

Everett, L.J., and Lazar, M.A. (2013). Cell-specific integration of nuclear receptor function at the genome. WIREs Systems Biology and Medicine *5*, 615–629.

Fang, B., Mane-Padros, D., Bolotin, E., Jiang, T., and Sladek, F.M. (2012). Identification of a binding motif specific to HNF4 by comparative analysis of multiple nuclear receptors. Nucleic Acids Research *40*, 5343–5356.

Farmer, S.R. (2008). Molecular determinants of brown adipocyte formation and function. Genes & Development *22*, 1269–1275.

Fu, M., Sun, T., Bookout, A.L., Downes, M., Yu, R.T., Evans, R.M., and Mangelsdorf, D.J. (2005). A Nuclear Receptor Atlas: 3T3-L1 Adipogenesis. Molecular Endocrinology *19*, 2437–2450.

Gampe, R.T., Jr., Montana, V.G., Lambert, M.H., Miller, A.B., Bledsoe, R.K., Milburn, M.V., Kliewer, S.A., Willson, T.M., and Xu, H.E. (2000). Asymmetry in the PPARγ/RXRα Crystal Structure Reveals the Molecular Basis of Heterodimerization among Nuclear Receptors. Molecular Cell *5*, 545–555.

Giguère, V., Hollenberg, S.M., Rosenfeld, M.G., and Evans, R.M. (1986). Functional domains of the human glucocorticoid receptor. Cell *46*, 645–652.

Giralt, M., and Villarroya, F. (2013). White, brown, beige/brite: different adipose cells for different functions? Endocrinology *154*, 2992–3000.

Glass, C.K., and Ogawa, S. (2005). Combinatorial roles of nuclear receptors in inflammation and immunity. Nature Reviews Immunology *6*, 44–55.

Golden, M.S., Cote, S.M., Sayeg, M., Zerbe, B.S., Villar, E.A., Beglov, D., Sazinsky, S.L., Georgiadis, R.M., Vajda, S., Kozakov, D., et al. (2013). Comprehensive Experimental and Computational Analysis of Binding Energy Hot Spots at the NF-κB Essential Modulator/IKKβ Protein–Protein Interface. Journal of the American Chemical Society *135*, 6242–6256.

Green, S., and Chambon, P. (1987). Oestradiol induction of a glucocorticoid-responsive gene by a chimaeric receptor. Nature *325*, 75–78.

Green, S., Walter, P., Kumar, V., Krust, A., Bornert, J.M., Argos, P., and Chambon, P. (1986). Human oestrogen receptor cDNA: sequence, expression and homology to v-erb-A. Nature *320*, 134–139.

Gronemeyer, H., and Bourguet, W. (2009). Allosteric Effects Govern Nuclear Receptor Action: DNA Appears as a Player. Science Signaling *2*, pe34–pe34.

Hall, J.M., McDonnell, D.P., and Korach, K.S. (2002). Allosteric regulation of estrogen receptor structure, function, and coactivator recruitment by different estrogen response elements. Molecular Endocrinology *16*, 469–486.

Heemers, H.V., and Tindall, D.J. (2010). Nuclear Receptor Coregulators: Promising Therapeutic Targets for the Treatment of Prostate Cancer. In Drug Management of Prostate Cancer, (New York, NY: Springer, New York, NY), pp. 41–51.

Heyman, R.A., Mangelsdorf, D.J., Dyck, J.A., Stein, R.B., Eichele, G., Evans, R.M., and Thaller, C. (1992). 9-cis retinoic acid is a high affinity ligand for the retinoid X receptor. Cell *68*, 397–406.

Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proceedings of the National Academy of Sciences USA *106*, 9362–9367.

Hollenberg, S.M., Weinberger, C., Ong, E.S., Cerelli, G., Oro, A., Lebo, R., Thompson, E.B., Rosenfeld, M.G., and Evans, R.M. (1985). Primary structure and expression of a functional human glucocorticoid receptor cDNA. Nature *318*, 635–641.

Hu, X., and Lazar, M.A. (1999). The CoRNR motif controls the recruitment of corepressors by nuclear hormone receptors. Nature *402*, 93–96.

Hudson, W.H., Youn, C., and Ortlund, E.A. (2012). The structural basis of direct glucocorticoid-mediated transrepression. Nature Structural and Molecular Biology *20*, 53–58.

Hurst, C.H., and Waxman, D.J. (2003). Activation of PPARalpha and PPARgamma by environmental phthalate monoesters. Toxicological Sciences *74*, 297–308.

IJpenberg, A., Jeannin, E., Wahli, W., and Desvergne, B. (1997). Polarity and Specific Sequence Requirements of Peroxisome Proliferator-activated Receptor (PPAR)/Retinoid X Receptor Heterodimer Binding to DNA. Journal of Biological Chemistry *272*, 20108–20117.

Isakova, A., Groux, R., Imbeault, M., Rainer, P., Alpern, D., Dainese, R., Ambrosini, G., Trono, D., Bucher, P., and Deplancke, B. (2017). SMiLE-seq identifies binding motifs of single and dimeric transcription factors. Nature Methods *14*, 316–322.

Issa, L.L., Leong, G.M., Barry, J.B., Sutherland, R.L., and Eisman, J.A. (2001). Glucocorticoid receptor-interacting protein-1 and receptor-associated coactivator-3 differentially interact with the vitamin D receptor (VDR) and regulate VDR-retinoid X receptor transcriptional cross-talk. Endocrinology *142*, 1606–1615.

Issemann, I., and Green, S. (1990). Activation of a member of the steroid hormone receptor superfamily by peroxisome proliferators. Nature *347*, 645–650.

Jacobi, D., Stanya, K., and Lee, C.-H. (2012). Adipose tissue signaling by nuclear receptors in metabolic complications of obesity. Adipocyte *1*, 4–12.

Jin, L., and Li, Y. (2010). Structural and functional insights into nuclear receptor signaling. Advanced Drug Delivery Reviews *62*, 1218–1226.

Johnson, A.B., and O'Malley, B.W. (2012). Steroid receptor coactivators 1, 2, and 3: critical regulators of nuclear receptor activity and steroid receptor modulator (SRM)-based cancer therapy. Molecular and Cellular Endocrinology *348*, 430–439.

Juge-Aubry, C., Pernin, A., Favez, T., Burger, A.G., Wahli, W., Meier, C.A., and Desvergne, B. (1997). DNA Binding Properties of Peroxisome Proliferator-activated Receptor Subtypes on Various Natural Peroxisome Proliferator Response Elements. Journal of Biological Chemistry *272*, 25252–25259.

Kassotis, C.D., Masse, L., Kim, S., Schlezinger, J.J., Webster, T.F., and Stapleton, H.M. (2017). Characterization of Adipogenic Chemicals in Three Different Cell Culture Systems: Implications for Reproducibility Based on Cell Source and Handling. Scientific Reports *7*, 42104.

Katz, R.W., Subauste, J.S., and Koenig, R.J. (1995). The Interplay of Half-site Sequence and Spacing on the Activity of Direct Repeat Thyroid Hormone Response Elements. Journal of Biological Chemistry *270*, 5238–5242.

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A., van der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S.R., Tan, G., et al. (2018). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Research *46*, D260–D266.

Kim, H.J., Yi, J.Y., Sung, H.S., Moore, D.D., Jhun, B.H., Lee, Y.C., and Lee, J.W. (1999). Activating signal cointegrator 1, a novel transcription coactivator of nuclear receptors, and its cytosolic localization under conditions of serum deprivation. Molecular and Cellular Biology *19*, 6323–6332.

Kliewer, S.A., Lehmann, J.M., and Willson, T.M. (1999). Orphan nuclear receptors: shifting endocrinology into reverse. Science *284*, 757–760.

Kliewer, S.A., Umesono, K., Noonan, D.J., Heyman, R.A., and Evans, R.M. (1992). Convergence of 9-cis retinoic acid and peroxisome proliferator signalling pathways through heterodimer formation of their receptors. Nature *358*, 771–774.

Koszewski, N.J., Malluche, H.H., and Russell, J. (2000). Vitamin D receptor interactions with positive and negative DNA response elements: an interference footprint comparison. Journal of Steroid Biochemistry and Molecular Biology *72*, 125–132.

Kulakovskiy, I.V., Medvedeva, Y.A., Schaefer, U., Kasianov, A.S., Vorontsov, I.E., Bajic, V.B., and Makeev, V.J. (2013). HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. Nucleic Acids Research *41*, D195–D202.

Kurokawa, R., Yu, V.C., Naar, A., Kyakumoto, S., Han, Z., Silverman, S., Rosenfeld, M.G., and Glass, C.K. (1993). Differential orientations of the DNA-binding domain and carboxy-terminal dimerization interface regulate binding site selection by nuclear receptor heterodimers. Genes & Development *7*, 1423–1435.

Kuzmič, P. (1996). Program DYNAFIT for the Analysis of Enzyme Kinetic Data: Application to HIV Proteinase. Analytical Biochemistry *237*, 260–273.

Le Zhan, Liu, H.-X., Fang, Y., Kong, B., He, Y., Zhong, X.-B., Fang, J., Wan, Y.-J.Y., and Guo, G.L. (2014). Genome-Wide Binding and Transcriptome Analysis of Human Farnesoid X Receptor in Primary Human Hepatocytes. PLoS ONE *9*, e105930.

Lefebvre, P., Mouchon, A., Lefebvre, B., and Formstecher, P. (1998). Binding of Retinoic Acid Receptor Heterodimers to DNA a role for histones NH2 termini. Journal of Biological Chemistry *273*, 12288–12295.

Lefterova, M.I., Steger, D.J., Zhuo, D., Qatanani, M., Mullican, S.E., Tuteja, G., Manduchi, E., Grant, G.R., and Lazar, M.A. (2010). Cell-Specific Determinants of Peroxisome Proliferator-Activated Receptor Function in Adipocytes and Macrophages. Molecular and Cellular Biology *30*, 2078–2089.

Lefterova, M.I., Zhang, Y., Steger, D.J., Schupp, M., Schug, J., Cristancho, A., Feng, D., Zhuo, D., Stoeckert, C.J., Liu, X.S., et al. (2008). PPAR and C/EBP factors orchestrate adipocyte biology via adjacent binding on a genome-wide scale. Genes & Development *22*, 2941–2952.

Lemay, D.G., and Hwang, D.H. (2006). Genome-wide identification of peroxisome proliferator response elements using integrated computational genomics. Journal of Lipids Research. *47*, 1583–1587.

Levin, A.A., Sturzenbecker, L.J., Kazmer, S., Bosakowski, T., Huselton, C., Allenby, G., Speck, J., Kratzeisen, C., Rosenberger, M., and Lovey, A. (1992). 9-cis retinoic acid stereoisomer binds and activates the nuclear receptor RXR alpha. Nature *355*, 359–361.

Lonard, D.M., and O'Malley, B.W. (2007). Nuclear Receptor Coregulators: Judges, Juries, and Executioners of Cellular Regulation. Molecular Cell *27*, 691–700.

Lou, X., Toresson, G., Benod, C., Suh, J.H., Philips, K.J., Webb, P., and Gustafsson, J.-Å. (2014). Structure of the retinoid X receptor α–liver X receptor β (RXRα–LXRβ) heterodimer on DNA. Nature Structural and Molecular Biology *21*, 277–281.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology *15*, 550.

Mader, S., Chen, J.Y., Chen, Z., White, J., Chambon, P., and Gronemeyer, H. (1993). The patterns of binding of RAR, RXR and TR homo- and heterodimers to direct repeats are dictated by the binding specificites of the DNA binding domains. The EMBO Journal *12*, 5029–5041.

Maloney, E.K., and Waxman, D.J. (1999). trans-Activation of PPARα and PPARγ by Structurally Diverse Environmental Chemicals. Toxicololgy and Applied Pharmacology *161*, 209–218.

Mangelsdorf, D.J., Ong, E.S., Dyck, J.A., and Evans, R.M. (1990). Nuclear receptor that identifies a novel retinoic acid response pathway. Nature *345*, 224–229.

Markov, G.V., Gutierrez-Mazariegos, J., Pitrat, D., Billas, I.M.L., Bonneton, F., Moras, D., Hasserodt, J., Lecointre, G., and Laudet, V. (2017). Origin of an ancient hormone/receptor couple revealed by resurrection of an ancestral estrogen. Science Advances *3*, e1601778.

Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., et al. (2006). TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Research *34*, D108–D110.

McInerney, E.M., Rose, D.W., Flynn, S.E., Westin, S., Mullen, T.M., Krones, A., Inostroza, J., Torchia, J., Nolte, R.T., Assa-Munt, N., et al. (1998). Determinants of coactivator LXXLL motif specificity in nuclear receptor transcriptional activation. Genes & Development *12*, 3357–3368.

Meier, C.A., Parkison, C., Chen, A., Ashizawa, K., Meier-Heusler, S.C., Muchmore, P., Cheng, S.Y., and Weintraub, B.D. (1993). Interaction of human beta 1 thyroid hormone receptor and its mutants with DNA and retinoid X receptor beta. T3 response element-dependent dominant negative potency. Journal of Clinical Investigation *92*, 1986–1993.

Meijsing, S.H., Pufall, M.A., So, A.Y., Bates, D.L., Chen, L., and Yamamoto, K.R. (2009). DNA Binding Site Sequence Directs Glucocorticoid Receptor Structure and Activity. Science *324*, 407–410.

Metzger, D., Imai, T., Jiang, M., Takukawa, R., Desvergne, B., Wahli, W., and Chambon, P. (2005). Functional role of RXRs and PPARgamma in mature adipocytes. Prostaglandins Leukotrienes Essential Fatty Acids *73*, 51–58.

Millard, C.J., Watson, P.J., Fairall, L., and Schwabe, J.W.R. (2013). An evolving understanding of nuclear receptor coregulator proteins. Journal of Molecular Endocrinology *51*, T23–T36.

Miyamoto, T., Kaneko, A., Kakizawa, T., Yajima, H., Kamijo, K., Sekine, R., Hiramatsu, K., Nishii, Y., Hashimoto, T., and Hashizume, K. (1997). Inhibition of Peroxisome Proliferator Signaling Pathways by Thyroid Hormone Receptor. Journal of Biological Chemistry *272*, 7752–7758.

Miyata, K.S., McCaw, S.E., Patel, H.V., Rachubinski, R.A., and Capone, J.P. (1996). The orphan nuclear hormone receptor LXR alpha interacts with the peroxisome proliferator-activated receptor and inhibits peroxisome proliferator signaling. Journal of Biological Chemistry *271*, 9189–9192.

Mohaghegh, N., Bray, D., Keenan, J., Penvose, A., Andrilenas, K.K., Ramlall, V., and Siggers, T. (2019). NextPBM: a platform to study cell-specific transcription factor binding and cooperativity. Nucleic Acids Research *47*, e31–e31.

Mossakowska, D.E. (1998). Expression of nuclear hormone receptors in Escherichia coli. Current Opinion in Biotechnology *9*, 502–505.

Näär, A.M., Boutin, J.-M., Lipkin, S.M., Yu, V.C., Holloway, J.M., Glass, C.K., and Rosenfeld, M.G. (1991). The orientation and spacing of core DNA-binding motifs dictate selective transcriptional responses to three nuclear receptors. Cell *65*, 1267–1279.

Nielsen, J. (2017). Systems Biology of Metabolism: A Driver for Developing Personalized and Precision Medicine. Cell Metabolism *25*, 572–579.

Nielsen, R., Pedersen, T.A., Hagenbeek, D., Moulos, P., Siersbaek, R., Megens, E., Denissov, S., Borgesen, M., Francoijs, K.J., Mandrup, S., et al. (2008). Genome-wide profiling of PPAR:RXR and RNA polymerase II occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis. Genes & Development *22*, 2953–2967.

Nomiyama, T., and Bruemmer, D. (2008). Liver X receptors as therapeutic targets in metabolism and atherosclerosis. Current Atherosclerosis Reports *10*, 88–95.

Palmer, C.N., Hsu, M.H., Griffin, H.J., and Johnson, E.F. (1995). Novel sequence determinants in peroxisome proliferator signaling. Journal of Biolical Chemistry *270*, 16114–16121.

Pap, A., Cuaranta-Monroy, I., Peloquin, M., and Nagy, L. (2016). Is the Mouse a Good Model of Human PPARγ-Related Metabolic Diseases? International Journal of Molecular Sciences *17*, 1236.

Penvose, A., Keenan, J.L., Bray, D., Ramlall, V., and Siggers, T. (2019). Comprehensive study of nuclear receptor DNA binding provides a revised framework for understanding receptor specificity. Nature Communications *10*, 2514.

Perissi, V., Staszewski, L.M., McInerney, E.M., Kurokawa, R., Krones, A., Rose, D.W., Lambert, M.H., Milburn, M.V., Glass, C.K., and Rosenfeld, M.G. (1999). Molecular determinants of nuclear receptor-corepressor interaction. Genes & Development *13*, 3198–3208.

Perissi, V., and Rosenfeld, M.G. (2005). Controlling nuclear receptors: the circular logic of cofactor cycles. Nature Reviews Molecular Cell Biology *6*, 542–554.

Perlmann, T., Rangarajan, P.N., Umesono, K., and Evans, R.M. (1993). Determinants for selective RAR and TR recognition of direct repeat HREs. Genes & Development *7*, 1411–1422.

Phan, T.Q., Jow, M.M., and Privalsky, M.L. (2010). DNA recognition by thyroid hormone and retinoic acid receptors: 3,4,5 rule modified. Molecular and Cellular Endocrinology *319*, 88–98.

Picard, F., Géhin, M., Annicotte, J.-S., Rocchi, S., Champy, M.-F., O'Malley, B.W., Chambon, P., and Auwerx, J. (2002). SRC-1 and TIF2 Control Energy Balance between White and Brown Adipose Tissues. Cell *111*, 931–941.

Pillai, H.K., Fang, M., Beglov, D., Kozakov, D., Vajda, S., Stapleton, H.M., Webster, T.F., and Schlezinger, J.J. (2014). Ligand Binding and Activation of PPARγ by Firemaster® 550: Effects on Adipogenesis and Osteogenesis in Vitro. Environmental Health Perspectives *122*, 1225–1232.

Poulard, C., Bittencourt, D., Wu, D.-Y., Hu, Y., Gerke, D.S., and Stallcup, M.R. (2017). A post-translational modification switch controls coactivator function of histone methyltransferases G9a and GLP. EMBO Reports *18*, 1442–1459.

Purcell, D.J., Jeong, K.W., Bittencourt, D., Gerke, D.S., and Stallcup, M.R. (2011). A distinct mechanism for coactivator versus corepressor function by histone methyltransferase G9a in transcriptional regulation. Journal of Biological Chemistry *286*, 41963–41971.

Qiang, L., Wang, L., Kon, N., Zhao, W., Lee, S., Zhang, Y., Rosenbaum, M., Zhao, Y., Gu, W., Farmer, S.R., et al. (2012). Brown remodeling of white adipose tissue by SirT1-dependent deacetylation of Pparγ. Cell *150*, 620–632.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Rastinejad, F., Huang, P., Chandra, V., and Khorasanizadeh, S. (2013). Understanding nuclear receptor form and function using structural biology. Journal of Molecular Endocrinology *51*, T1–T21.

Rastinejad, F., Perlmann, T., Evans, R.M., and Sigler, P.B. (1995). Structural determinants of nuclear receptor assembly on DNA direct repeats. Nature *375*, 203–211.

Rosen, E.D., Sarraf, P., Troy, A.E., Bradwin, G., Moore, K., Milstone, D.S., Spiegelman, B.M., and Mortensen, R.M. (1999). PPAR gamma is required for the differentiation of adipose tissue in vivo and in vitro. Molecular Cell *4*, 611–617.

Rosenfeld, M.G., and Glass, C.K. (2001). Coregulator Codes of Transcriptional Regulation by Nuclear Receptors. Journal of Biological Chemistry *276*, 36865–36868.

Sap, J., Muñoz, A., Damm, K., Goldberg, Y., Ghysdael, J., Leutz, A., Beug, H., and Vennström, B. (1986). The c- erb -A protein is a high-affinity receptor for thyroid hormone. Nature *324*, 635–640.

Savic, D., Ramaker, R.C., Roberts, B.S., Dean, E.C., Burwell, T.C., Meadows, S.K., Cooper, S.J., Garabedian, M.J., Gertz, J., and Myers, R.M. (2016). Distinct gene regulatory programs define the inhibitory effects of liver X receptors and PPARG on cancer cell proliferation. Genome Medicine 2016 8:1 *8*, 74.

Schöne, S., Bothe, M., Einfeldt, E., Borschiwer, M., Benner, P., Vingron, M., Thomas-Chollier, M., and Meijsing, S.H. (2018). Synthetic STARR-seq reveals how DNA shape and sequence modulate transcriptional output and noise. PLoS Genetics *14*, e1007793.

Schöne, S., Jurk, M., Helabad, M.B., Dror, I., Lebars, I., Kieffer, B., Imhof, P., Rohs, R., Vingron, M., Thomas-Chollier, M., et al. (2016). Sequences flanking the core-binding site modulate glucocorticoid receptor structure and activity. Nature Communications *7*, 12621.

Schwabe, J.W.R., Chapman, L., Finch, J.T., and Rhodes, D. (1993). The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: How receptors discriminate between their response elements. Cell *75*, 567–578.

Seo, J.B., Moon, H.M., Kim, W.S., Lee, Y.S., Jeong, H.W., Yoo, E.J., Ham, J., Kang, H., Park, M.-G., Steffensen, K.R., et al. (2004). Activated liver X receptors stimulate adipocyte differentiation through induction of peroxisome proliferator-activated receptor gamma expression. Molecular and Cellular Biology *24*, 3430–3444.

Siersbaek, M.S., Loft, A., Aagaard, M.M., Nielsen, R., Schmidt, S.F., Petrovic, N., Nedergaard, J., and Mandrup, S. (2012). Genome-Wide Profiling of Peroxisome Proliferator-Activated Receptor in Primary Epididymal, Inguinal, and Brown Adipocytes Reveals Depot-Selective Binding Correlated with Gene Expression. Molecular and Cellular Biology *32*, 3452–3463.

Siggers, T., Chang, A.B., Teixeira, A., Wong, D., Williams, K.J., Ahmed, B., Ragoussis, J., Udalova, I.A., Smale, S.T., and Bulyk, M.L. (2011). Principles of dimer-specific gene regulation revealed by a comprehensive characterization of NF-κB family DNA binding. Nature Immunology *13*, 95–102.

Smale, S.T. (2001). Core promoters: active contributors to combinatorial gene regulation. Genes & Development *15*, 2503–2508.

Soccio, R.E., Chen, E.R., Rajapurkar, S.R., Safabakhsh, P., Marinis, J.M., Dispirito, J.R., Emmett, M.J., Briggs, E.R., Bin Fang, Everett, L.J., et al. (2015). Genetic Variation Determines PPARγ Function and Anti-diabetic Drug Response In Vivo. *162*, 33–44.

Spivakov, M. (2014). Spurious transcription factor binding: Non-functional or genetically redundant? BioEssays *36*, 798–806.

Stormo, G.D., and Zhao, Y. (2010). Determining the specificity of protein–DNA interactions. Nature Reviews Genetics *11*, 751–760.

Structural Genomics Consortium, China Structural Genomics Consortium, Northeast Structural Genomics Consortium, Gräslund, S., Nordlund, P., Weigelt, J., Hallberg, B.M., Bray, J., Gileadi, O., Knapp, S., et al. (2008). Protein production and purification. Nature Methods *5*, 135–146.

Surjit, M., Ganti, K.P., Mukherji, A., Ye, T., Hua, G., Metzger, D., Li, M., and Chambon, P. (2011). Widespread Negative Response Elements Mediate Direct Repression by Agonist- Liganded Glucocorticoid Receptor. Cell *145*, 224–241.

Temple, K.A., Cohen, R.N., Wondisford, S.R., Yu, C., Deplewski, D., and Wondisford, F.E. (2005). An Intact DNA-binding Domain Is Not Required for Peroxisome Proliferator-activated Receptor γ (PPARγ) Binding and Activation on Some PPAR Response Elements. Journal of Biological Chemistry *280*, 3529–3540.

Tenbaum, S., and Baniahmad, A. (1997). Nuclear receptors: Structure, function and involvement in disease. The International Journal of Biochemistry & Cell Biology *29*, 1325–1341.

Tobin, K.A., Steineger, H.H., Alberti, S., Spydevold, O., Auwerx, J., Gustafsson, J.A., and Nebb, H.I. (2000). Cross-talk between fatty acid and cholesterol metabolism mediated by liver X receptor-alpha. Molecular Endocrinology *14*, 741–752.

Treuter, E., Albrektsen, T., Johansson, L., Leers, J., and Gustafsson, J.A. (1998). A regulatory role for RIP140 in nuclear receptor activation. Molecular Endocrinology *12*, 864–881.

Tung, E.W.Y., Ahmed, S., Peshdary, V., and Atlas, E. (2017). Firemaster® 550 and its components isopropylated triphenyl phosphate and triphenyl phosphate enhance adipogenesis and transcriptional activity of peroxisome proliferator activated receptor (Pparγ) on the adipocyte protein 2 (aP2) promoter. PLoS ONE *12*, e0175855.

Umesono, K., Murakami, K.K., Thompson, C.C., and Evans, R.M. (1991). Direct repeats as selective response elements for the thyroid hormone, retinoic acid, and vitamin D3 receptors. Cell *65*, 1255–1266.

Uppenberg, J., Svensson, C., Jaki, M., Bertilsson, G., Jendeberg, L., and Berkenstam, A. (1998). Crystal structure of the ligand binding domain of the human nuclear receptor PPARgamma. Journal of Biological Chemistry *273*, 31108–31112.

Villanueva, C.J., Vergnes, L., Wang, J., Drew, B.G., Hong, C., Tu, Y., Hu, Y., Peng, X., Xu, F., Saez, E., et al. (2013). Adipose Subtype-Selective Recruitment of TLE3 or Prdm16 by PPARγ Specifies Lipid Storage versus Thermogenic Gene Programs. Cell Metabolism *17*, 423–435.

Viswakarma, N., Jia, Y., Bai, L., Vluggens, A., Borensztajn, J., Xu, J., and Reddy, J.K. (2010). Coactivators in PPAR-Regulated Gene Expression. PPAR Research *2010*, 1–21.

Wang, D., Zhu, W., Chen, L., Yan, J., Teng, M., and Zhou, Z. (2018a). Neonatal triphenyl phosphate and its metabolite diphenyl phosphate exposure induce sex- and dose-dependent metabolic disruptions in adult mice. Environmental Pollution *237*, 10–17.

Wang, N., Zou, Q., Xu, J., Zhang, J., and Liu, J. (2018b). Ligand binding and heterodimerization with retinoid X receptor α (RXRα) induce farnesoid X receptor (FXR) conformational changes affecting coactivator binding. Journal of Biological Chemistry *293*, 18180–18191.

Wang, V.Y.-F., Huang, W., Asagiri, M., Spann, N., Hoffmann, A., Glass, C., and Ghosh, G. (2012). The Transcriptional Specificity of NF-κB Dimers Is Coded within the κB DNA Response Elements. Cell Reports *2*, 824–839.

Watson, L.C., Kuchenbecker, K.M., Schiller, B.J., Gross, J.D., Pufall, M.A., and Yamamoto, K.R. (2013). The glucocorticoid receptor dimer interface allosterically transmits sequence-specific DNA signals. Nature Structural and Molecular Biology *20*, 876–883.

Weikum, E.R., Liu, X., and Ortlund, E.A. (2018). The nuclear receptor superfamily: A structural perspective. Protein Science. *27*, 1876–1892.

Weikum, E.R., de Vera, I.M.S., Nwachukwu, J.C., Hudson, W.H., Nettles, K.W., Kojetin, D.J., and Ortlund, E.A. (2017). Tethering not required: the glucocorticoid receptor binds directly to activator protein-1 recognition motifs to repress inflammatory genes. Nucleic Acids Research *45*, 8596–8608.

Weinberger, C., Hollenberg, S.M., Rosenfeld, M.G., and Evans, R.M. (1985). Domain structure of human glucocorticoid receptor and its relationship to the v-erb-A oncogene product. Nature *318*, 670–672.

Willson, T.M., and Moore, J.T. (2002). Genomics versus orphan nuclear receptors--a half-time report. Molecular Endocrinology *16*, 1135–1144.

Wurtz, J.-M., Bourguet, W., Renaud, J.-P., Vivat, V., Chambon, P., Moras, D., and Gronemeyer, H. (1996). A canonical structure for the ligand-binding domain of nuclear receptors. Nature Structural and Molecular Biology *3*, 87–94.

Xie, W., Barwick, J.L., Simon, C.M., Pierce, A.M., Safe, S., Blumberg, B., Guzelian, P.S., and Evans, R.M. (2000). Reciprocal activation of xenobiotic response genes by nuclear receptors SXR/PXR and CAR. Genes & Development *14*, 3014–3023.

Yang, L., Zhou, T., Dror, I., Mathelier, A., Wasserman, W.W., Gordân, R., and Rohs, R. (2014). TFBSshape: a motif database for DNA shape features of transcription factor binding sites. Nucleic Acids Research *42*, D148–D155.

Zechel, C., Shen, X.Q., Chambon, P., and Gronemeyer, H. (1994). Dimerization interfaces formed between the DNA binding domains determine the cooperative binding of RXR/RAR and RXR/TR heterodimers to DR5 and DR4 elements. The EMBO Journal *13*, 1414–1424.

Zhang, X.-K., Hoffmann, B., Tran, P.B.V., Graupner, G., and Pfahl, M. (1992). Retinoid X receptor is an auxiliary protein for thyroid hormone and retinoic acid receptors. Nature *355*, 441–446.

Zhang, X.M., and Verdine, G.L. (1999). A Small Region in HMG I(Y) Is Critical for Cooperation with NF-κB on DNA. Journal of Biological Chemistry *274*, 20235–20243.

Zhou, T., Shen, N., Yang, L., Abe, N., Horton, J., Mann, R.S., Bussemaker, H.J., Gordân, R., and Rohs, R. (2015). Quantitative modeling of transcription factor binding specificities using DNA shape. Proceedings of the National Academy of Sciences USA *112*, 4654–4659.

CURRICULUM VITAE